DOCUMENT RESUME

ED 322 213 TM 015 458

| | |
|---|---|
| AUTHOR | Samejima, Fumiko |
| TITLE | Content-Based Observation of Informative Distractors and Efficiency of Ability Estimation. |
| INSTITUTION | Tennessee Univ., Knoxville. Dept. of Psychology. |
| SPONS AGENCY | Office of Naval Research, Arlington, Va. |
| REPORT NO | ONR-RR-90-05 |
| PUB DATE | 5 Jun 90 |
| CONTRACT | N00014-87-K-0320-4421-549 |
| NOTE | 54p. |
| PUB TYPE | Reports - Evaluative/Feasibility (142) |
| EDRS PRICE | MF01/PC03 Plus Postage. |
| DESCRIPTORS | *Ability Identification; *Distractors (Tests); Equations (Mathematics); *Estimation (Mathematics); *Item Response Theory; Mathematical Models; *Multiple Choice Tests; Objective Tests; Test Construction; Testing Problems; Test Interpretation; Test Items; *Test Use |
| IDENTIFIERS | *Ability Estimates; Graded Response Model; Nominal Response Model |

ABSTRACT

The shortcomings of the conventional way of using and interpreting multiple-choice tests are summarized. Some theories and methodologies that can be applied for better use multiple-choice test items are described. Empirical facts are introduced to support the theoretical observations. New strategies are proposed that will reduce "noise" and lead to more efficient ability estimation. In so doing, simple-minded avoidance of non-monotonicity of the operating characteristic of the correct answer is reconsidered, and the efficient use of informative distractors is taken into account. This paper deals with nominal response and multiple-choice test situations in which examinees are required to choose one of the given alternatives in connection with the graded response model. The operating characteristic of a given response pattern is a product of the operating characteristics of the item response categories contained in that response pattern. The multiple-choice test situation should be treated as a special instance of the nominal level of response, although the problem of random or irrational choice should be investigated. Twenty-four graphs are included. (SLD)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

ONR/RR-90-05

# CONTENT-BASED OBSERVATION OF INFORMATIVE DISTRACTORS AND EFFICIENCY OF ABILITY ESTIMATION

FUMIKO SAMEJIMA

UNIVERSITY OF TENNESSEE

KNOXVILLE, TENN. 37996-0900

JUNE, 1990

Prepared under the contract number N00014-87-K-0327
(RR-04-08 prog. No.
Cognitive Science Research Program
Cognitive and Neural Sciences Division
Office of Naval Research

Approved for public release; distribution unlimited.
Reproduction in whole or in part is permitted for
any purpose of the United States Government.

Best Copy Available

BEST COPY AVAILABLE

SECURITY CLASSIFICATION OF THIS PAGE

## REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

| | |
|---|---|
| 1a REPORT SECURITY CLASSIFICATION: Unclassified | 1b RESTRICTIVE MARKINGS |
| 2a SECURITY CLASSIFICATION AUTHORITY | 3 DISTRIBUTION / AVAILABILITY OF REPORT: Approved for public release; Distribution unlimited |
| 2b DECLASSIFICATION / DOWNGRADING SCHEDULE | |
| 4 PERFORMING ORGANIZATION REPORT NUMBER(S) | 5 MONITORING ORGANIZATION REPORT NUMBER(S) |
| 6a NAME OF PERFORMING ORGANIZATION: Fumiko Samejima, Ph.D. Psychology Department; 6b OFFICE SYMBOL (If applicable) | 7a NAME OF MONITORING ORGANIZATION: Cognitive Science 1142 CS |
| 6c ADDRESS (City, State, and ZIP Code): 310B Austin Peay Building The University of Tennessee Knoxville, TN 37996-0900 | 7b ADDRESS (City, State, and ZIP Code): Office of Naval Research 800 N. Quincy Street Arlington, VA 22217 |
| 8a NAME OF FUNDING / SPONSORING ORGANIZATION: Cognitive Science Research Program; 8b OFFICE SYMBOL (If applicable) | 9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: N00014-87-K-0320 |
| 8c ADDRESS (City, State, and ZIP Code): Office of Naval Research 800 N. Quincy Street Arlington, VA 22217 | 10 SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO 61153N; PROJECT NO RR-042-04; TASK NO 042-04-01; WORK UNIT ACCESSION NO 4421-549 |

11 TITLE (Include Security Classification): Content-based observation of informative distractors and efficiency of ability estimation

12 PERSONAL AUTHOR(S): Fumiko Samejima, Ph.D.

| 13a TYPE OF REPORT | 13b TIME COVERED | 14 DATE OF REPORT (Year, Month, Day) | 15 PAGE COUNT |
|---|---|---|---|
| technical report | FROM 1987 TO 1990 | June 5, 1990 | 50 |

16 SUPPLEMENTARY NOTATION

| 17 COSATI CODES | | | 18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number) |
|---|---|---|---|
| FIELD | GROUP | SUB-GROUP | Latent Trait Models, Multiple-Choice Test, Effect of Noise, Test Item Writing |
| | | | |
| | | | |

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

The paper summarises the shortages of the conventional way of handling the multiple-choice test and also describes theories and methodologies that can be applied for a better handling of the multiple-choice test items; some empirical facts are introduced to support the theoretical observations; finally new strategies of item writing are proposed which will reduce *noise* and lead to more efficient ability estimation. In so doing, simple-minded avoidance of non-monotonicity of the operating characteristic of the correct answer is reconsidered, and efficient use of informative distractors is taken into account.

| 20 DISTRIBUTION / AVAILABILITY OF ABSTRACT | 21 ABSTRACT SECURITY CLASSIFICATION | |
|---|---|---|
| ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ DTIC USERS | | |
| 22a NAME OF RESPONSIBLE INDIVIDUAL: Dr. Charles E. Davis | 22b TELEPHONE (Include Area Code): 202-696-4046 | 22c OFFICE SYMBOL: ONR-1142-CS |

DD Form 1473, JUN 86 — Previous editions are obsolete. — SECURITY CLASSIFICATION OF THIS PAGE

S/N 0102-LF-014-6603

## TABLE OF CONTENTS


| | | Page |
|---|---|---|
| 1 | Introduction | 1 |
| 2 | Non-Monotonicity of the Conditional Probability of the Positive Response, Given Latent Variable | 1 |
| 3 | Effect of Noise in the Three-Parameter Logistic Model and the Meanings of the Difficulty and Discrimination Parameters | 9 |
| 4 | Informative Distractors of the Multiple-Choice Test Item | 14 |
| 5 | Merits of the Nonparametric Approach for the Identification of Informative Distractors and for the Estimation of the Operating Characteristics of an Item | 18 |
| 6 | Efficiency in Ability Estimation and Strategies of Writing Test Items | 28 |
| 7 | Discussion and Conclusions | 39 |

REFERENCES

The research was conducted at the principal investigator's laboratory, 405 Austin Peay Bldg., Department of Psychology, University of Tennessee, Knoxville, Tennessee. Those who worked as assistants for this research include Christine A. Golik, Barbara A. Livingston, Lee Hai Gan and Nancy H. Domm.

# I Introduction

As early as in 1968 the author wrote and discussed the conceivable non-monotonicity of the operating characteristic of the correct answer of the multiple-choice test item, which is based strictly upon theory (cf. Samejima, 1968). Since then, such a phenomenon has actually been observed with empirical data. For example, Lord and Novick reported such a curve when they plotted the percent of the correct answer against the test score for each item as an approximation to the item characteristic function (cf. Lord and Novick, 1968, Chapter 16). Since, as their Theorem 16.4.1 states, *the average, over all items, of the sample item-test regressions falls along a straight line through the origin with forty-five degree slope,* such a *dip* cannot be detected for an easy item even if it exists, as far as we use the item-test regression as an approximation. It is quite possible, therefore, that there are more than one item among those items that have such "dips"; only they were not detected.

In recent years, several more direct approaches of estimating operating characteristics have revealed such "dips" among ASVAB items. Partly because of the availability of computer software, such as Logist (Wingersky, Barton and Lord, 1982), Bilog (Bock and Atkin, 1981), etc., however, it is a common procedure among researchers that they *mold* non-monotonic operating characteristics of correct answers into the three-parameter logistic model, ignoring the non-monotonicity. In some cases, even strategies are taken so that distractors, which cause the non-monotonicity, are considered as undesirable ones and are replaced by some other *non-threatening* alternative answers.

A question must be raised as to whether this strategy is wise. In the present paper, this issue will be discussed both from theory and from practice, and a new strategy of writing test items, which leads to more efficient ability estimation, will be proposed. It will take advantage of the ease in handling mathematics attributed to parameterization, and yet minimize the effect of noise caused by random guessing.

# II Non-Monotonicity of the Conditional Probability of the Positive Response, Given Latent Variable

This section is basically the essence or a summary of the paper published by the author more than twenty years ago (Samejima, 1968), as one of the research reports of the L. L. Thurstone Psychometric Laboratory of the University of North Carolina. The content of the paper was a protocol which led to the proposal of a *new* family of models for the multiple-choice test item (Samejima, 1979b). The author believes that this paper published in 1968 still gives *new ideas* to today's research communities.

The paper deals with the nominal response, and also multiple-choice situations, in which examinees are required to choose one of the given alternatives, in connection with the graded response model (cf. Samejima, 1969, 1972). Let $\theta$ denote the latent variable, or ability, which assumes real numbers. Let $g\ (= 1, 2, ..., n)$ denote an item, $k_g$ be a discrete response to item $g$, and $P_{k_g}(\theta)$ denote the operating characteristic of the discrete response $k_g$, or the conditional probability, given $\theta$, with which the examinee responds to item $g$ with $k_g$. Throughout the paper the principle of *local independence* is assumed to be valid, so that within any group of examinees all characterized by the same value of the latent variable $\theta$ the distributions of the item response categories are all independent of each other. Thus the operating characteristic of a given response pattern is a product of the operating characteristics of the item response categories contained in that response pattern (cf. Lord and Novick, 1968).

For a multiple-choice item a certain number of false answers are given in addition to the correct answer. In a general case it is impossible to score them in a graded manner in accordance with their degrees of attainment toward the goal. Thus the multiple-choice situation should be treated as a special instance of the nominal level of response, although, in addition, the problem of random or irrational choice should be investigated.

Confining discussions to examinees who have responded to item $g$ incorrectly, there can be diversity of false answers if they have responded to it freely, without being forced to choose one of a set of alternative answers. It is conceivable that some of the false answers may require high levels of ability measured while some others may not, some may be related to the ability measured strongly while some others may not, etc. An objective measure of the plausibility of a specified false answer is its operating characteristic, i.e., the probability of its occurrence defined for a fixed value of ability $\theta$, and, therefore, expressed as a function of $\theta$.

Let $M_s(\theta)$ be a sequence of the conditional probabilities corresponding to the cognitive subprocesses required in finding the plausibility of response $k_g$ to item $g$, and $U_{k_g}(\theta)$ be the conditional probability that an examinee discovers the irrationality of response $k_g$ as the answer to item $g$, on condition that he has already found out its plausibility. The operating characteristic of $k_g$, which is denoted by $P_{k_g}(\theta)$, can be expressed by

$$P_{k_g}(\theta) = [1 - U_{k_g}(\theta)] \prod_{s \in k_g} M_s(\theta) \quad , \tag{2.1}$$

since it is reasonably assumed that an examinee who gives a response $k_g$ to item $g$ is one who has succeeded in finding $k_g$'s plausibility, and yet failed in finding its irrationality. We notice that this formula is exactly the same in its structure as the definition of $P_{x_g}(\theta)$ on the graded response level, where $M_s(\theta)$ is replaced by $M_i(\theta)$ and $U_{k_g}(\theta)$ is replaced by $M_{(x_g+1)}(\theta)$ (cf. Samejima, 1972). Defining $M_{k_g}(\theta)$ such that

$$M_{k_g}(\theta) = \prod_{s \in k_g} M_s(\theta) \quad , \tag{2.2}$$

we can rewrite (2.1) into

$$P_{k_g}(\theta) = M_{k_g}(\theta)[1 - U_{k_g}(\theta)] \quad . \tag{2.3}$$

It will reasonably be assumed from their definitions that both $M_{k_g}(\theta)$ and $U_{k_g}(\theta)$ be strictly increasing in $\theta$, provided that a specified response $k_g$ is a *good mistake* in the sense that the discoveries of its plausibility and irrationality are properly related with ability $\theta$. It will also be reasonably assumed that the upper asymptotes of $M_{k_g}(\theta)$ and $U_{k_g}(\theta)$ are unity, and the lower asymptote of $M_{k_g}(\theta)$ is zero.

We assume that both $M_{k_g}(\theta)$ and $U_{k_g}(\theta)$ are three-times-differentiable with respect to $\theta$. It is easily observed that, in order to satisfy the unique maximum condition (Samejima, 1969, 1972), $P_{k_g}(\theta)$ defined by (2.3) must fulfill the following inequalities:

$$\frac{\partial^2}{\partial\theta^2} \log M_{k_g}(\theta) = \frac{\partial}{\partial\theta}\left[\frac{\partial}{\partial\theta} M_{k_g}(\theta)\{M_{k_g}(\theta)\}^{-1}\right] < 0 \tag{2.4}$$

and

$$\frac{\partial^2}{\partial\theta^2} \log[1 - U_{k_g}(\theta)] = \frac{\partial}{\partial\theta}\left[-\frac{\partial}{\partial\theta} U_{k_g}(\theta)\{1 - U_{k_g}(\theta)\}^{-1}\right] < 0 \quad . \tag{2.5}$$

(For proof, see Samejima, 1968.) Note that in this case the lower asymptote of $U_{k_g}(\theta)$ need not be zero. The operating characteristic of a specified response $k_g$ which satisfies the unique maximum condition was called the *plausibility curve* (Samejima, 1968), and later the *plausibility function* (cf. Samejima,

1984a). As the condition suggests, the plausibility curve is necessarily unimodal. A schematized hypothesis for the plausibility curve will be the following. The probability that an examinee will find the plausibility, but will *fail* in discovering the irrationality, of a specified response $k_g$ as the answer to item $g$ is a function of ability $\theta$ ; it increases as ability $\theta$ increases, reaches maximum at a certain value of $\theta$ , and then decreases afterwards. If an item provides many such responses, their plausibility curves will be powerful sources of information in estimating examinees' abilities. That is to say, we can make use of specific wrong answers to an item as sources of information, as well as the correct answer.

Let $P_g(\theta)$ denote the operating characteristic of the correct answer of a dichotomous item $g$ in the free-response situation. Let $P_g^*(\theta)$ be the same function, but in the multiple-choice situation. The conventional three-parameter model is represented by

$$P_g^*(\theta) = c_g + (1 - c_g)P_g(\theta) \quad , \tag{2.6}$$

where $c_g$ is the probability with which an examinee will guess correctly (Lord and Novick, 1968). This is a monotonically increasing function of $\theta$ with $c_g$ $(\geq 0)$ and unity as its lower and upper asymptotes, provided that $P_g(\theta)$ is strictly increasing in $\theta$ with zero and unity as its lower and upper asymptotes.

The psychological hypothesis which has led to the formula (2.6) in the multiple-choice situation is the following. If an examinee has ability $\theta$ , then the probability that he will know the correct answer is given by $P_g(\theta)$ ; if he does not know it, he will guess randomly, and, with probability $c_g$ , will guess correctly (Lord and Novick, 1968). Thus we have for the operating characteristic of the correct answer of item $g$ in the multiple-choice situation

$$P_g(\theta) + [1 - P_g(\theta)]c_g \quad , \tag{2.7}$$

which leads to (2.6). This hypothesis may not necessarily be appropriate for ability measurement. One can never tell in the measurement of a reasoning ability, for instance, whether an examinee *knows* the correct answer to item $g$ or not, until he has tried to solve it. He may respond with an incorrect alternative without guessing at all. To explain such a case we need some other hypothesis than the one which leads to the formula (2.6).

Hereafter, we assume that $P_g(\theta)$ is *strictly increasing* in $\theta$ with zero and unity as its lower and upper asymptotes, and is twice-differentiable with respect to $\theta$ . Suppose, further, that both $P_g(\theta)$ and $[1 - P_g(\theta)]$ satisfy the unique maximum condition. In this case $P_g^*(\theta)$ defined by (2.6) *does not* satisfy either of Conditions (i) and (ii) for the unique maximum, unless $c_g$ is zero, i.e., the free-response situation, although they are fulfilled for the negative answer to item $g$ (cf. Samejima, 1973). Observations and discussion are made (Samejima, 1968) giving two simple cases of the multiple-choice situation as examples. In those examples, only two items are involved, and the response pattern, (1,0), is solely treated, and precise mathematical derivations are given.

A possible correction for the conventional functional formula for the operating characteristic of the correct answer of a multiple-choice item can be made by introducing the probability of random guessing defined for a fixed value of $\theta$ . Let $d_g(\theta)$ denote this probability. A reasonable assumption for this function may be that it be *non-increasing* in $\theta$ . Thus the probability with which an examinee of ability $\theta$ will answer item $g$ correctly by following the due cognitive process is expressed by $[1 - d_g(\theta)]P_g(\theta)$ ; and the one with which he will give the correct answer by guessing should be $d_g(\theta)c_g$ . For economy of notation, let $P_g^*(\theta)$ be the operating characteristic of the correct answer to item $g$ in the corrected functional formula also. We can write

$$P_g^*(\theta) \;=\; [1 - d_g(\theta)]P_g(\theta) \;+\; d_g(\theta)c_g \tag{2.8}$$

$$= P_g(\theta) + d_g(\theta)[c_g - P_g(\theta)] \ .$$

A schematised psychological hypothesis which leads to this formula is as the following. If an examinee has ability $\theta$, then he will depend upon random guessing in answering item $g$ with probability $d_g(\theta)$; in that case, the conditional probability with which he will guess correctly is given by $c_g$. If he does not depend upon random guessing, he will try to solve the item by the due cognitive process, and will succeed in solving it with probability $P_g(\theta)$. Thus according to this functional formula the probability with which an examinee will respond with an incorrect alternative without guessing is given by $[1 - d_g(\theta)][1 - P_g(\theta)]$, which is nil in the model represented by the formula (2.6).

We can conceive of several factors which may affect the functional formula for $d_g(\theta)$. The difficulty of item $g$ may be one of them; the discriminating power may be another; the number of alternatives attached to item $g$ may also affect the probability, i.e., it may be that the fewer the number of alternatives, the more tempted to depend upon random guessing an examinee will be; also the plausibilities of the alternatives may be counted as a factor.

In a simplified case where $d_g(\theta)$ is constant throughout the whole range of $\theta$, we can rewrite (2.8) in the following form.

$$P_g^*(\theta) = d_g c_g + [1 - d_g]P_g(\theta) \ . \tag{2.9}$$

This is somewhat similar to formula (2.6), the conventional functional formula for the operating characteristic of the correct answer of a multiple-choice item. The lower asymptote of the present function is $d_g c_g$ $(\leq c_g)$, however, while it is $c_g$ in (2.6); the upper asymptote of the present function is $[1 - d_g(1 - c_g)]$, *which can be less than unity*, while it is unity in (2.6). In a special case where $d_g = 0$, that is, an examinee tries to solve item $g$ by proper reasoning with probability one, (2.9) reduces to $P_g(\theta)$, the operating characteristic of the correct answer in the free-response situation. In another special case where $d_g = 1$, that is, an examinee depends upon random guessing with probability one, (2.9) reduces to a constant, $c_g$. In the more general case where $d_g(\theta)$ varies as $\theta$ varies, it is observed from (2.8) that

$$\begin{cases} 0 < P_g(\theta) \leq P_g^*(\theta) \leq c_g & ; \text{if} \quad \theta < \theta_0 \\ P_g^*(\theta) = c_g = P_g(\theta) & ; \text{if} \quad \theta = \theta_0 \\ c_g \leq P_g^*(\theta) \leq P_g(\theta) < 1 & ; \text{if} \quad \theta > \theta_0 \end{cases} \tag{2.10}$$

where

$$\theta_0 = P_g^{-1}(c_g) \ , \tag{2.11}$$

provided that $c_g$ is greater than zero. This result is quite natural, since it is reasonably assumed that the probability of success in solving item $g$ will decrease by random guessing if the one attained by the due cognitive process is higher than the one attained by random guessing, and it will increase by random guessing if the latter probability is higher than the former. If we assume that the asymptotes of $d_g(\theta)$ in negative and positive directions be unity and zero, respectively, we will obtain $c_g$ and unity as the lower and upper asymptotes of $P_g^*(\theta)$. Figure 2-1 presents two examples of the operating characteristic given by (2.8) where $c_g$ is 0.2, using two different $d_g(\theta)$'s. Note there is a "dip" on the lower part of the curves for $P_g^*(\theta)$. These two $d_g(\theta)$'s are identical for the lower levels of $\theta$,

but differ on the upper levels, with the upper asymptotes 0.0 and 0.1, respectively. In these examples, therefore, the upper asymptote of $P_g^*(\theta)$ is unity in the first example, and 0.92 in the second, i.e., the conditional probability for the correct answer never approaches unity however high the ability may be.

If $d_g(\theta)$ is differentiable, $P_g^*(\theta)$ is also differentiable, and from (2.8) we have

$$\frac{\partial}{\partial\theta}P_g^*(\theta) = [1 - d_g(\theta)]\frac{\partial}{\partial\theta}P_g(\theta) + [c_g - P_g(\theta)]\frac{\partial}{\partial\theta}d_g(\theta) \quad . \tag{2.12}$$

Thus it is obvious that $P_g^*(\theta)$ is strictly increasing in $\theta$ for the range $\theta \geq \theta_0$, if, and only if, $d_g(\theta)$ is less than unity for the range of $\theta$ satisfying $\theta \geq \theta_0$. Thus in this case $P_g^*(\theta)$ is non-decreasing in $\theta$ throughout its whole range. In general, $P_g^*(\theta)$ equals $c_g$ and presents a horizontal line as far as $d_g(\theta)$ is unity, and then increases for the rest of the range as $\theta$ increases.

As for the range expressed by $\theta \leq \theta_0$, $P_g^*(\theta)$ equals $c_g$ regardless of the value of $P_g(\theta)$ for the values of $\theta$ for which $d_g(\theta)$ is unity, and is some positive value less than $c_g$ otherwise. If $d_g(\theta)$ is unity throughout this range of $\theta$, $P_g^*(\theta)$ presents a horizontal line for this range. If $d_g(\theta)$ is unity for the negative extreme value of $\theta$, but $d_g(\theta)$ takes on some values less than unity for a subset of $\theta$ of this range, $P_g^*(\theta)$ has at least one local minimum. If $d_g(\theta)$ is less than unity for the negative extreme value of $\theta$, $P_g^*(\theta)$ can be strictly increasing in $\theta$, non-decreasing, or have *one or more local minima*, in accordance with the functional formula for $d_g(\theta)$.

It is obvious that any operating characteristic having local minima does not satisfy the unique maximum condition (Samejima, 1969, 1972), and neither does the one whose first derivative equals zero at some value of $\theta$. In the case of $P_g^*(\theta)$ defined by (2.8) we can prove that, in general, it *does not* satisfy the unique maximum condition, even if it is strictly increasing in $\theta$. (For proof, see Samejima, 1968.)

Two characteristics of the model represented by (2.8) are that it allows "dips", and also a smaller value than unity for the upper asymptote of the operating characteristic of the correct answer, as Figure 2-1 illustrates. In these examples, there is only one "dip" on the lower level of $\theta$. There can be more than one, however, and an example is presented elsewhere (Samejima, 1968). In many cases the model may describe the real operating characteristic of the correct answer more closely than the three-parameter model.

It has been reported by several researchers that they have come across estimated operating characteristics of correct answers that do not converge to unity, but to some other values less than unity. Note that the general model described above can handle such situations, although most of the other models proposed by different researchers so far cannot.

We notice that neither (2.2) nor (2.8) explicitly takes into consideration the influences of separate distractors. Suppose an examinee A has chosen to solve item $g$ by reasoning, i.e., without guessing, and has reached an answer which is not correct. Suppose, further, that this specified response is *not* given as an alternative answer to this item. Then either he will decide to give an answer by guessing, or he will try to solve the item by reasoning all over again. To account for these possibilities, we would have to give practically all the different plausible responses to item $g$ as its alternatives, which is practically impossible, since the number of alternative answers is more or less restricted. In contrast to this, it is interesting to note that the psychological hypothesis behind the three-parameter logistic model may be more realistic in the case where no very plausible responses except for the correct answer to item $g$ are given as its alternative answers. Thus, even if an examinee has reached a specified plausible response other than the correct answer, he may turn to random guessing simply because he cannot find that specified answer among the alternatives. Such a situation has another serious problem, however, since it is likely for an examinee who is highly alternative-oriented to choose the correct answer without much reasoning or guessing, simply because the other alternatives are too ridiculous to be the answer to the item. As the result, the operating characteristic of the correct answer may be *deformed* so that

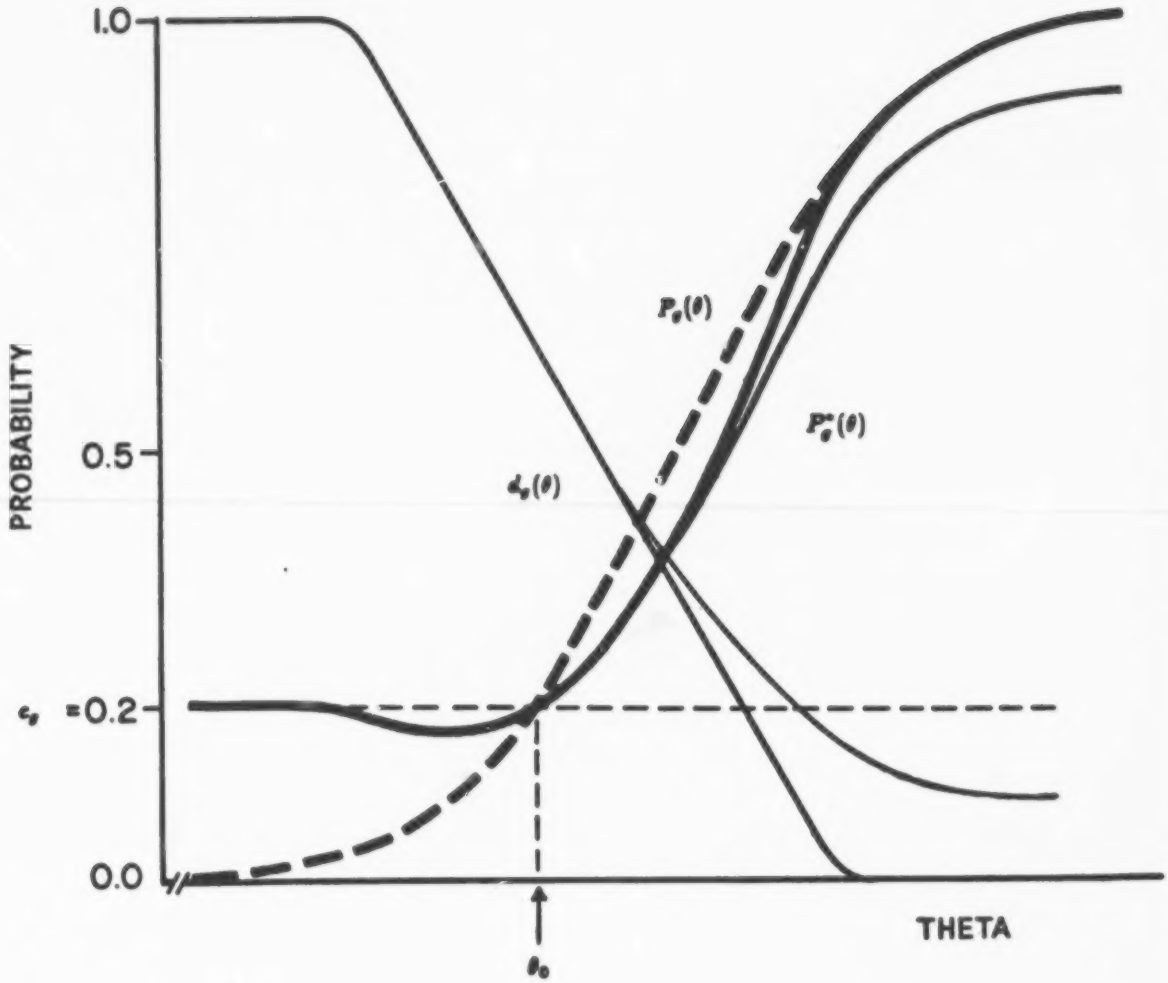



**FIGURE 2-1**

Relationships among $P_g(\theta)$, $d_g(\theta)$ and $P_g^*(\theta)$ Using Two Different $d_g(\theta)$'s.

it has a lower difficulty and less discriminating power. Plausible answers as *distractors* are necessary as alternatives in order not to destroy the nature of the item.

It is conceivable that the plausibilities of the alternatives attached to item $g$ other than the correct answer will be one of the factors affecting the probability of random guessing in the multiple-choice situation. As distinct from the discussion developed in the preceding section, here we shall suppose that an examinee will try to solve the item following proper cognitive processes at the beginning, and only in the case where he has reached an answer which is not given as an alternative, or where he has failed to find any answer at all, he will guess.

Let $k_g$ or $h_g$ denote a specified response to item $g$ which is given as an alternative, including the correct answer, and $P_{k_g}(\theta)$ or $P_{h_g}(\theta)$ be its operating characteristic in the free-response situation. It may reasonably be assumed that $\sum_{k_g} P_{k_g}(\theta)$ is less than or equal to unity for any fixed value of $\theta$. Let $P^*_{k_g}(\theta)$ or $P^*_{h_g}(\theta)$ denote the operating characteristic of a specified alternative $k_g$ or $h_g$ in the multiple-choice situation, and $c_{k_g}$ or $c_{h_g}$ be the probability of choosing $k_g$ or $h_g$ by guessing, which satisfies

$$\sum_{k_g} c_{k_g} = 1 \ . \tag{2.13}$$

Thus we can write

$$P^*_{k_g}(\theta) = P_{k_g}(\theta) + [1 - \sum_{h_g} P_{h_g}(\theta)]\, c_{k_g} \tag{2.14}$$

for any $k_g$, and, by using the notation for the correct answer as we did in the previous sections, we obtain

$$P^*_g(\theta) = P_g(\theta) + [1 - \sum_{h_g} P_{h_g}(\theta)]\, c_g \ . \tag{2.15}$$

It is worth noting that we have specified not only the operating characteristic of the correct answer in the multiple-choice situation, but also of each distractor. The utility of the operating characteristic of each wrong alternative answer in the estimation of an examinee's ability, as well as the one of the correct response, is suggested, and this is a feature of the present discussion.

It has been made clear that, in general, $P^*_g(\theta)$ *does not* satisfy the unique maximum condition regardless of the functional formulae for the plausibility curves of the distractors. As for the alternatives other than the correct answer, it can easily be shown that, in general, $P^*_{k_g}(\theta)$ does not satisfy the unique maximum condition (cf. Samejima, 1968, 1979b).

For the purpose of illustration, Figure 2-2 presents a simple example in which only two alternatives, the correct answer and one incorrect response, are given. In this example, $P^*_{k_g}(\theta)$ for the wrong answer is drawn from the ceiling in order to make the picture visibly understandable. A normal ogive function given by

$$P_g(\theta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_g(\theta - b_g)} \exp\{-u^2/2\}\, du \tag{2.16}$$

where $a_g = 1/1.48$ and $b_g = 0.36$ is used as the operating characteristic of the correct answer, and the same formula is applied for $U_{k_g}(\theta)$ and $M_{k_g}(\theta)$ for the incorrect response. The corresponding values of parameters are 1/1.23 and -1.84 for $M_{k_g}(\theta)$, and 1/1.51 and -0.83 for $U_{k_g}(\theta)$. The value of $c_g$, as well as that of $c_{k_g}$ for the incorrect answer, is 0.5.

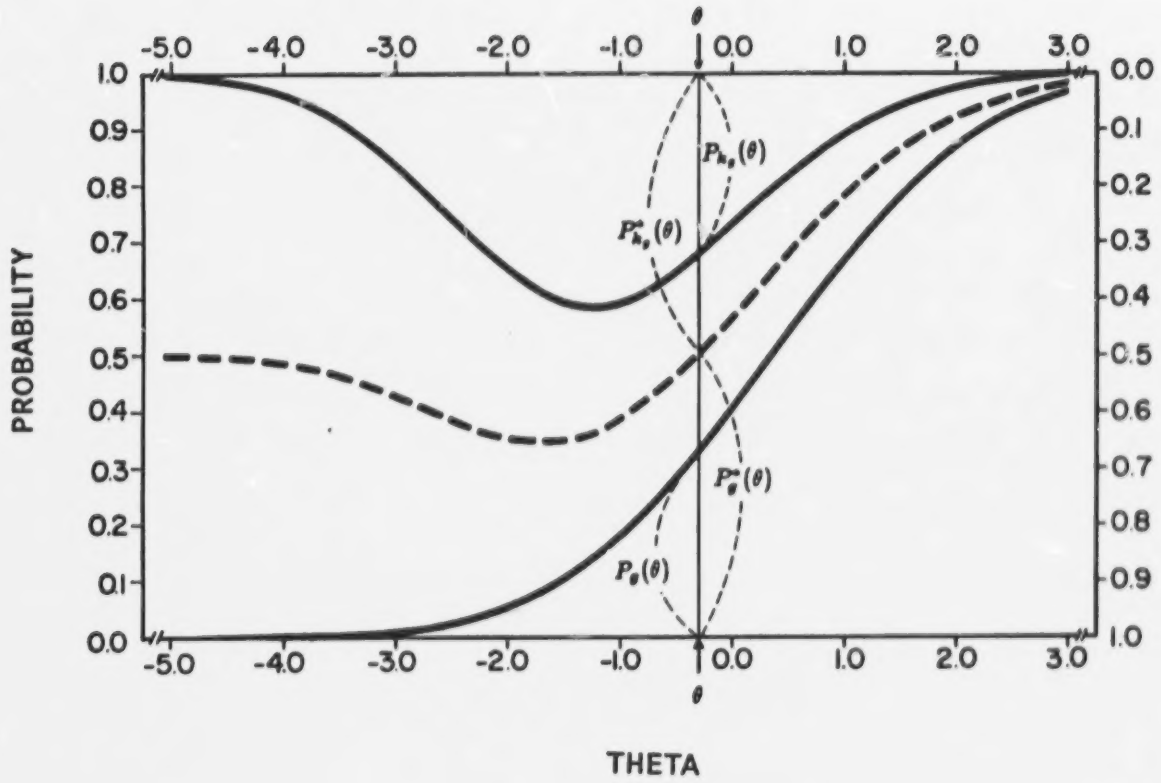



FIGURE 2-2

**Operating Characteristic of the Correct Answer in the Free-Response Situation (Solid Line) and in the Multiple-Choice Situation (Dashed Line), in the Case Where Only Two Alternatives Are Given; Also the Operating Characteristic of the Other Alternative in the Free-Response Situation (Solid Line) is Plotted from the Ceiling; $c_g = c_{h_g} = 0.5$ .**

It is obvious from the above observations and discussion that these are the fundamental philosophies which led to the proposal of the "new" family of models for the multiple-choice test item (Samejima, 1979b). These philosophies will provide us with the idea of content-based observation of informative distractors and strategies of writing test items, which will be proposed in a later section. The general model described here is called Informative Distractor Model, in contrast with the Equivalent Distractor Model, to which the three-parameter model represented by (2.6) belongs (cf. Samejima, 1979b).

# III Effect of Noise in the Three-Parameter Logistic Model and the Meanings of the Difficulty and Discrimination Parameters

Three-parameter logistic model for the multiple-choice test item is represented by

$$P_g^*(\theta) = c_g + (1 - c_g)[1 + \exp\{-Da_g(\theta - b_g)\}]^{-1} \quad , \tag{3.1}$$

where $a_g$ , $b_g$ , and $c_g$ are the item discrimination, difficulty, and guessing parameters, and $D$ is a scaling factor, which is set equal to 1.7 when the logistic model is used as a substitute for the normal ogive model. When $c_g = 0$ , this formula provide us with the operating characteristic of the correct answer in the original logistic model.

It is still a common procedure among researchers to adopt the three-parameter logistic model for their multiple-choice test items and compare the resulting estimated discrimination parameters, or the difficulty parameters, across different items. An important fact that is overlooked is that *this is not legitimate*, for the addition of the third parameter $c_g$ makes the other two item parameters lose their original meanings. If $a_g = 1.00$ and $c_g = 0.25$ in the three-parameter logistic model, for example, this corresponds to $a_g = 0.75$ in the logistic model in the maximum discrimination power. If, in addition to these parameter values, $b_g = 0.00$ , then the difficulty level for the three-parameter logistic model defined as the level of $\theta$ at which chances for success are 0.5 is -0.4077336 , i.e., substantially lower than 0.00 .

In general, we can write

$$\begin{cases} \alpha_g = (1 - c_g)\, a_g \\ \beta_g = b_g + (Da_g)^{-1} \log (1 - 2c_g) \quad , \end{cases} \tag{3.2}$$

where $\alpha_g$ denotes the actual discrimination power and $\beta_g$ is the actual difficulty level in the three-parameter logistic model. As we can see in (3.2), the effect of the third parameter $c_g$ can be substantial, both on the discrimination power $\alpha_g$ and on the difficulty index $\beta_g$ . Thus the simple comparison of the values of $a_g$ for two or more test items having different values of the lower asymptote $c_g$ is illegitimate and can be harmful, for the factor $(1 - c_g)$ may affect the value of $\alpha_g$ , the real discrimination power, substantially. As for the difficulty index, since the second term on the right hand side of the second equation of (3.2) is always negative for $0 < c_g < 0.5$ , this term represents the amount of decrement of the difficulty level. Note that as $c_g$ tends to 0.5 , $\beta_g$ approaches negative infinity! (If $c_g \geq 0.5$ then $\beta_g$ does not even exist.) The illegitimacy of, and the danger in, comparing $b_g$'s across two or more test items having different lower asymptotes $c_g$ is even more obvious for the difficulty index.

Figure 3-1 presents the operating characteristic of the correct answer in the normal ogive model with $a_g = 1.00$ and $b_g = 0.00$ by a dotted line, the one in the logistic model with the same parameters and the scaling factor, $D = 1.7$ , by a solid line, and the one in the three-parameter logistic model with

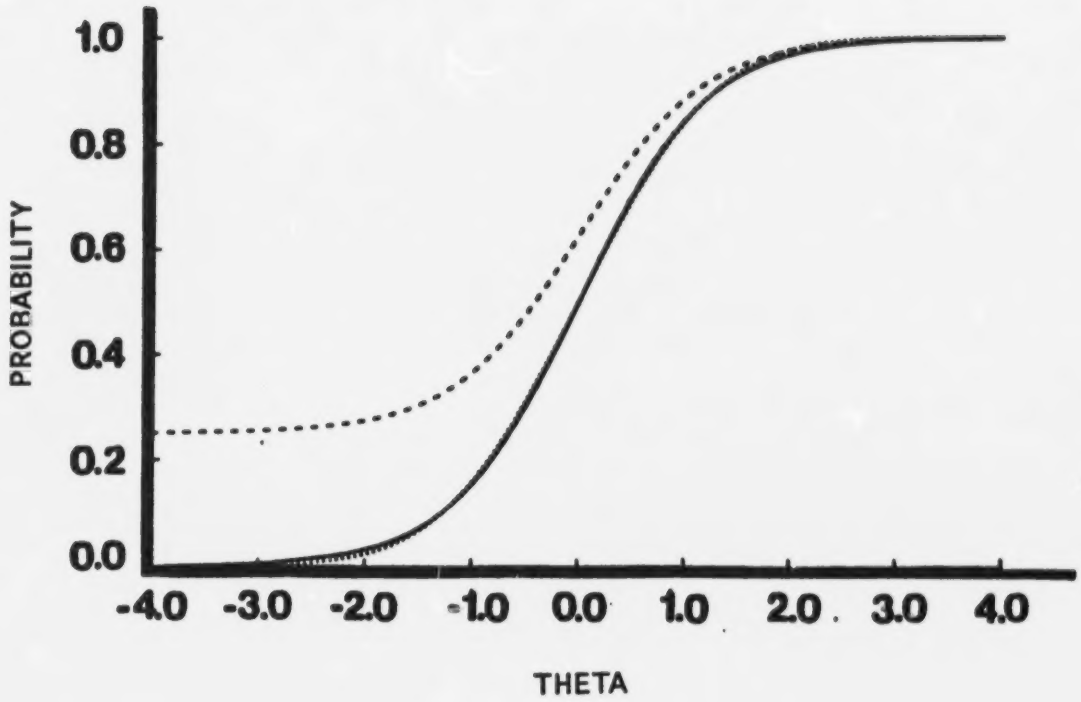



FIGURE 3-1

Operating Characteristics in the Normal Ogive Model (Dotted Line), in the Logistic Model (Solid Line) and in the Three-Parameter Logistic Model (Dashed Line), with the Parameters $a_g = 1.0$, $b_g = 0.0$, $c_g = 0.25$ and the Scaling Factor $D = 1.7$.

the same two item parameters and scaling factor and the third parameter, $c_g = 0.25$ , by a dashed line. It is obvious from theory that for all the three operating characteristics of the correct answer the derivatives are highest at $\theta = b_g = 0.0$ . Actually, these three derivatives are: $(2\pi)^{-1/2}a_g$ , $Da_g/4$ and $(1-c_g)Da_g/4$ , respectively, for the three functions in Figure 3-1. The ratio of this maximal slope in the normal ogive model to the one in the logistic model is approximately 0.938687718 , which is not so much less than unity. The corresponding ratio between the three-parameter logistic model and the logistic model is $(1-c_g)$ , which equals 0.75 when $c_g = 0.25$ , and is as low as 0.50 when $c_g = 0.50$ . The ratio between the three-parameter logistic model and the normal ogive model is approximately $0.938687718(1-c_g)$ , which is a little less than $(1-c_g)$ .

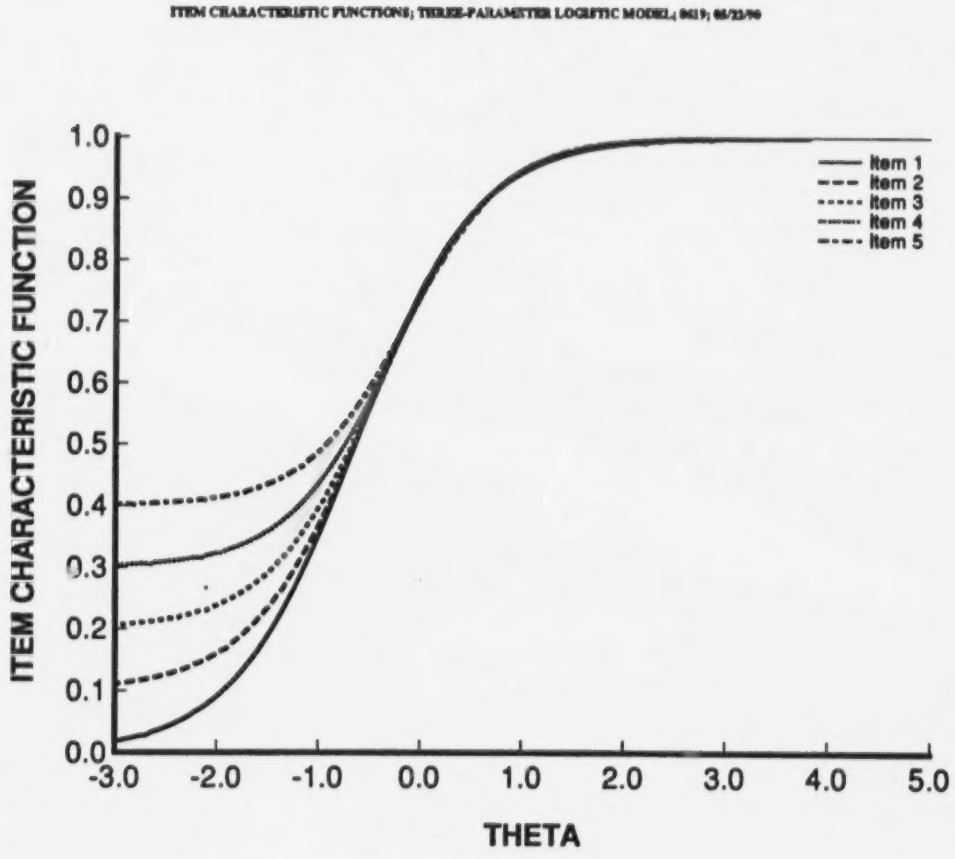
Figure 3-2 illustrates that several sets of substantially different parameter values in the three-parameter logistic model can produce very similar operating characteristics of the correct answer. We can tell that the differences in the values of the discrimination and difficulty parameters for these items are substantial, and yet the resulting curves are very close to each other for a wide range of $\theta$ . *Simple comparison of the two estimated discrimination parameters is illegitimate*, therefore, when the estimated guessing parameters proved to be different from each other, as is usually the case with actual data. Since the estimation of the third parameter $c_g$ tends to be most inaccurate, this example indicates the danger in direct comparisons of the estimated discrimination parameters, and also the estimated difficulty parameters, across the items.

In most cases the estimated guessing parameter of a multiple-choice test item provides us with some other value than the reciprocal of the number of the alternative answers. It is reported that in some cases the estimated $c_g$ takes on quite high values (cf. Lord, 1980, Section 2.2). These phenomena suggest that the philosophy behind the model is unrealistic. Researchers using the three-parameter logistic model argue, however, that it still is a convenient approximation to real operating characteristics of correct answers, because of its simplicity in mathematics. In a way it is true. The effective use of the three-parameter model cannot be realised, however, unless we know the problems attributed to the model, and use the model in such a way that these weaknesses will not cause too much noise and inefficiency.

Investigation of the problems encountered when we apply the three-parameter logistic model to the data which actually follow the normal ogive model was made earlier (Samejima, 1984b). The data used in the study are simulated data for two samples of 500 and 2,000 hypothetical examinees, respectively, sampled from the uniform ability distribution for the interval of $\theta$ , (-2.5, 2.5). In order to investigate the effect of the number of test items on the resultant estimated parameters obtained by Logist 5, we used: 1) Ten Item Test and 2) Thirty-Five Item Test, both of which consist of binary items following the normal ogive model. The response pattern for each hypothetical subject was produced by Monte Carlo Method. Combining these two hypothetical tests, we observed the result of: 3) Forty-Five Item Test, and, in addition, we observed the result of rather artificially created: 4) Eighty Item Test (cf. Samejima, 1984b).

These results suggest that there exists a substantial effect of the assumed third parameter, $c_g$ , on the other two estimated item parameters, if the estimation is made by *molding* the operating characteristic of the correct answer into that of the three-parameter logistic model, when actually it follows the normal ogive model. This effect appears to be stronger on the estimated discrimination parameter than on the estimated difficulty parameter. In order to amend these enhancements, the discrimination shrinkage factor and the difficulty reduction index were proposed (Samejima, 1984b) by formulae (3.4) and (3.6), respectively.

(3.3) $$a_g^* = \varsigma(c_g^*)\, a_g \ .$$




## FIGURE 3-2

**Examples of the Operating Characteristics of the Correct Answer in the Three-Parameter Logistic Model (Dotted Lines), Together with the One in the Logistic Model with $a_g = 1.00$ and $b_g = -0.64$ (Solid Line). The Parameters for the Four Functions in the Order of $a_g$, $b_g$ and $c_g$ are: 1.05, -0.52, 0.10; 1.10, -0.40, 0.20; 1.15, -0.27, 0.30; 1.20, -0.13, 0.40; respectively.**

(3.4) $$\zeta(c_g^*) = -\log(1-2c_g^*)\log(1+c_g^*) - \log(1-c_g^*)^{-1} \ .$$

(3.5) $$b_g^* = b_g + \xi(c_g^* \mid a_g) \ .$$

(3.6) $$\xi(c_g^* \mid a_g) = (Da_g)^{-1}\log(1+c_g^*) - \log(1-c_g^*) \ .$$

In these formulae, $a_g^*$ , $b_g^*$ , and $c_g^*$ indicate the estimated item discrimination, difficulty and guessing parameters when the three-parameter logistic model is assumed, respectively. Some resulting estimated operating characteristics of the correct answer turned out to be disastrously different from the theorectical functions, especially when only ten binary test items were included. We find no substantial differences between the results of 500 Subject Case and 2,000 Subject Case, indicating that increasing the number of subjects from 500 to 2,000 does not provide us with a substantial gain.

It has been pointed out that the three-parameter logistic model does not satisfy the unique maximum condition for the likelihood function, and this topic has been thoroughly discussed (Samejima, 1973). The expected loss of item information for a fixed value of $\theta$ is given by

(3.7) $$I_g(\theta) - I_g^*(\theta) = c_g D^2 a_g^2 [\{\psi_g(\theta)\}^2\{1-\psi_g(\theta)\}][c_g + (1-c_g)\psi_g(\theta)]^{-1} \ ,$$

where

(3.8) $$\psi_g(\theta) = [1+\exp\{-Da_g((\theta)-b_g)\}]^{-1} \ ,$$

and $I_g(\theta)$ and $I_g^*(\theta)$ are the item information functions in the logistic and the three-parameter logistic models, respectively. We have for the critical value $\ell_g$ , below which the information provided by the correct answer to the item following the three-parameter logistic model assumes negative values

(3.9) $$\ell_g = b_g + (2Da_g)^{-1}\log c_g \ ,$$

which is strictly increasing with the increase in the parameter value $c_g$ , and also in $a_g$ and in $b_g$ . If, for example, $a_g = 1.00$ and $b_g = 0.00$ , $\ell_g = -0.473364$ for $c_g = 0.20$ , and $\ell_g = -0.407734$ for $c_g = 0.25$ . They are considerably high values relative to $b_g$ .

An important implication is that $\ell_g$ is the point of $\theta$ below which the existence of a unique maximum likelihood estimate is not assured for all the response patterns which include the correct answer to item $g$ . Although this warning has been ignored by most researchers for many years, a recent research (Yen, Burket and Sykes, in press) points out this is happening much more often than people might think.

It has been pointed out (Samejima, 1979a, 1982a) that there is a certain constancy in the total amount of item information, regardless of the parameter values and of specific functional formulae for the operating characteristic of the correct answer. If, for example, the model belongs to Type A, i.e., the operating characteristic of the correct answer is monotone increasing with zero and unity as its lower and upper asymptotes, respectively, then the total area under the curve of the square root of the item information function will equal $\pi$ . If the model belongs to Type B, i.e., the same as Type A

except that the lower asymptote of the operating characteristic of the correct answer is greater than zero, as is the case with the three-parameter logistic model, then the total area will become

$$\pi - 2\tan^{-1}[c_g(1-c_g)^{-1}]^{1/2} \quad , \tag{3.10}$$

with the second and last term as *the loss in the amount of total item information.* This last term is strictly a function of $c_g$ . When $c_g = 0.20$ , for example, the total amount of item information reduces, approximately, to $0.705\pi$ , and when $c_g = 0.25$ it is approximately equal to $0.667\pi$ . More observations concerning the effect of noise in the three-parameter logistic model have been made elsewhere (Samejima, 1982b).

As all the above observations indicate, the addition of the third parameter, $c_g$ , to the logistic model creates many negative results. We have seen that these negative effects are greater for larger values of $c_g$ . In using the three-parameter logistic model as an approximation to real operating characteristics, therefore, we need to take these facts into consideration. Among others, if we are in a situation where we can modify or revise our items, we must try to reduce the effect of noise coming from $c_g$ as much as possible. Strategies of writing the multiple-choice test items must be considered accordingly.

# IV Informative Distractors of the Multiple-Choice Test Item

So far most observations and discussion have been focued on theory. Applications of certain non-parametric methods of estimating the operating characteristics for some empirical data have revealed, however, that many multiple-choice test items do not follow the three-parameter model, nor do they follow the Equivalent Distractor Model in general, to which the three-parameter logistic model belongs. Those items can best be interpreted by the Informative Distractor Model.

Figure 4-1 presents two examples of the set of operating characteristics of the four alternative answers to an item taken from the Level 11 Vocabulary Test of the Iowa Test of Basic Skills, which were estimated by the Simple Sum Procedure of the Conditional P.D.F. Approach combined with the Normal Approach Method (Samejima, 1981). We can see in these graphs that each distractor has its own unique operating characteristic, or *plausibility function*, and also the estimated operating characteristic of the correct answer is fairly close to the one in the normal ogive model, which is drawn by a solid line in the figure. This set of operating characteristics can better be represented by one of the family of models proposed for the multiple-choice test item, which was originated by the philosophy described in Section 2 and takes account of the unique information provided by each distractor as well as the effect of the examinee's random guessing behavior (cf. Samejima, 1979b). Figure 4-2 illustrates the operating characteristic of the correct answer in Model A. We can see that it is very close to the one in the normal ogive model which is drawn by a dotted line, except for the lower part of the curve, the conditional probability of success which is almost entirely caused by random guessing. In cases like this, it will be wise to approximate the curve by the normal ogive function by discarding the item response in estimating lower ability, since it provides us with nothing but noise, as was discussed in the preceding section.

Detailed observations for the plausibility functions of *distractors* are made elsewhere (Samejima, 1984a), for the forty-three items of the Level 11 Vocabulary Subtest of the Iowa Test of Basic Skills. Similar discoveries have also been reported with respect to many ASVAB test items. In those results, it is clear that separate wrong answers given as alternatives provide us with differential informations. As long as we adopt models like the three-parameter logistic model, however, we will never *discover* nor can we make use of these differential informations, which can be useful in ability estimation in the sense that they will substantially increase the accuracy of estimation.

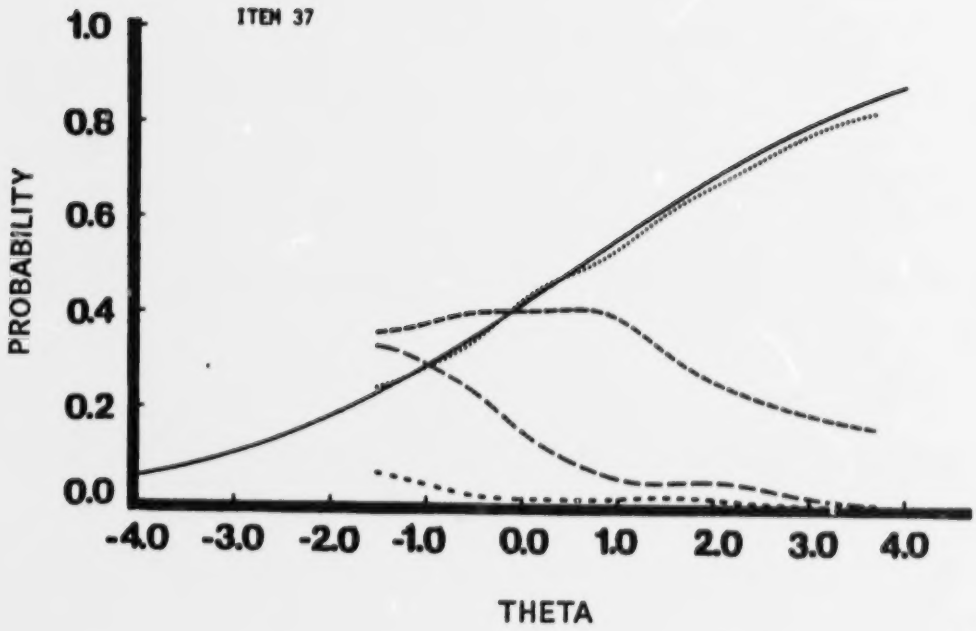



FIGURE 4-1

Two Examples of the Estimated Operating Characteristics of the Correct Answer (Dotted Line) and of the Three Distractors (Dashed Lines) Obtained by the Simple Sum Procedure of the Conditional P.D.F. Approach Combined with the Normal Approach Method Together with the One for the Correct Answer Obtained by Assuming the Normal Ogive Model (Solid Line) Taken from the Level 11 Vocabulary Subtest of the Iowa Test of Basic Skills.

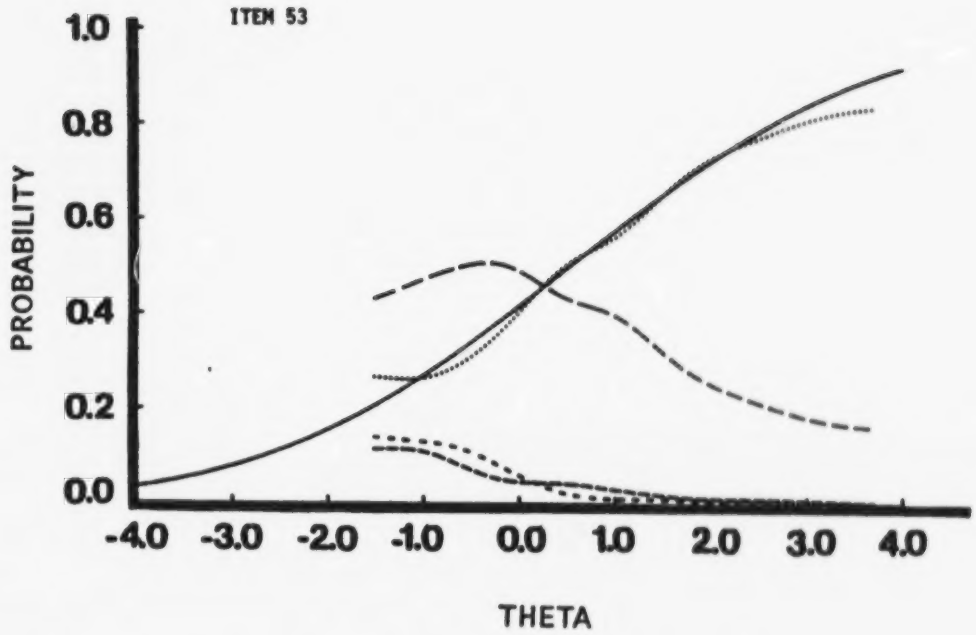



FIGURE 4-1 (Continued)

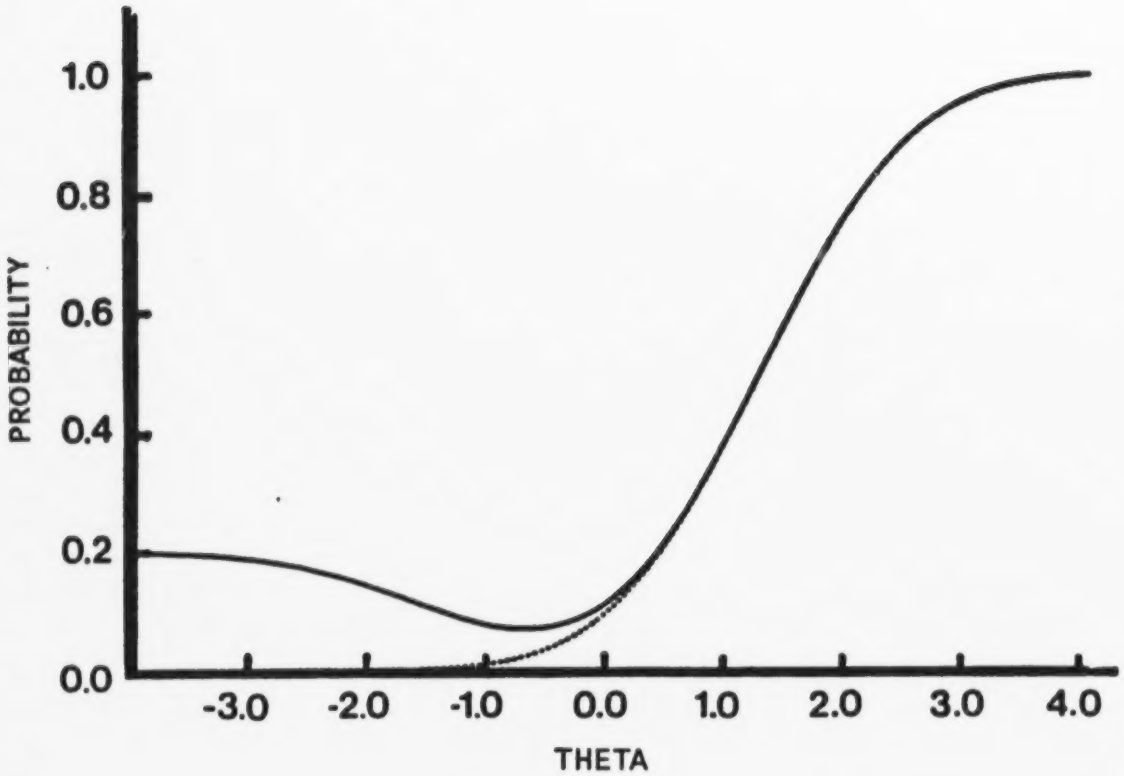



FIGURE 4-2

Example of the Operating Characteristic of the Correct Answer in Model A (Solid Line) Together with One in the Normal Ogive Model (Dotted Line).

How can we approach the plausibility functions of distractors? One way of handling this issue may be to adopt a model that belongs to the family of models (Samejima, 1979b) described earlier, and estimate the parameters involved in the model. Another more scientific approach may be to use a nonparametric method of estimating the operating characteristic. In this way, we shall be able to *discover* the plausibility functions of our distractors, rather than *mold* them into some mathematical formula. This will be discussed in the subsequent section, proposing a new approach to item analysis. Practical suggestions in writing items, which take all these facts and observations into account, will be given in Section 6.

# V Merits of the Nonparametric Approach for the Identification of Informative Distractors and for the Estimation of the Operating Characteristics of an Item

Methods and approaches developed for estimating the operating characteristics of discrete item responses without assuming any mathematical form (cf. Samejima, 1981, 1990) enable us to find out whether or not a given incorrect alternative answer to a multiple-choice test item is informative in the sense that it contributes to the increment in the accuracy in the estimation of the individual's ability. In the past years various sets of data based upon the Vocabulary Subtest of the Iowa Tests of Basic Skills, upon Shiba's Word/Phrase Comprehension Tests, ASVAB Tests of Word Knowledge and of Math Knowledge, etc., have been analyzed by using, mainly, the Simple Sum Procedure of Conditional P.D.F. Approach combined with the Normal Approach Method (cf. Samejima, 1981). Recently, the author proposed a new approach, which is called Differential Weight Procedure of the Conditional P.D.F. Approach. This new method has been introduced and discussed in a separate paper (Samejima, 1990). Here we shall only introduce the essence of the method, and illustrate some results on simulated data.

The item response information function, $I_{k_g}(\theta)$, is defined by

$$I_{k_g}(\theta) = -\frac{\partial^2}{\partial\theta^2}\log\ P_{k_g}(\theta)\ . \tag{5.1}$$

The item information function $I_g(\theta)$ is the conditional expectation of the item response information function, given $\theta$, so that we can write

$$I_g(\theta) = E[I_{k_g}(\theta)\mid\theta] = \sum_{k_g} I_{k_g}(\theta)P_{k_g}(\theta)\ . \tag{5.2}$$

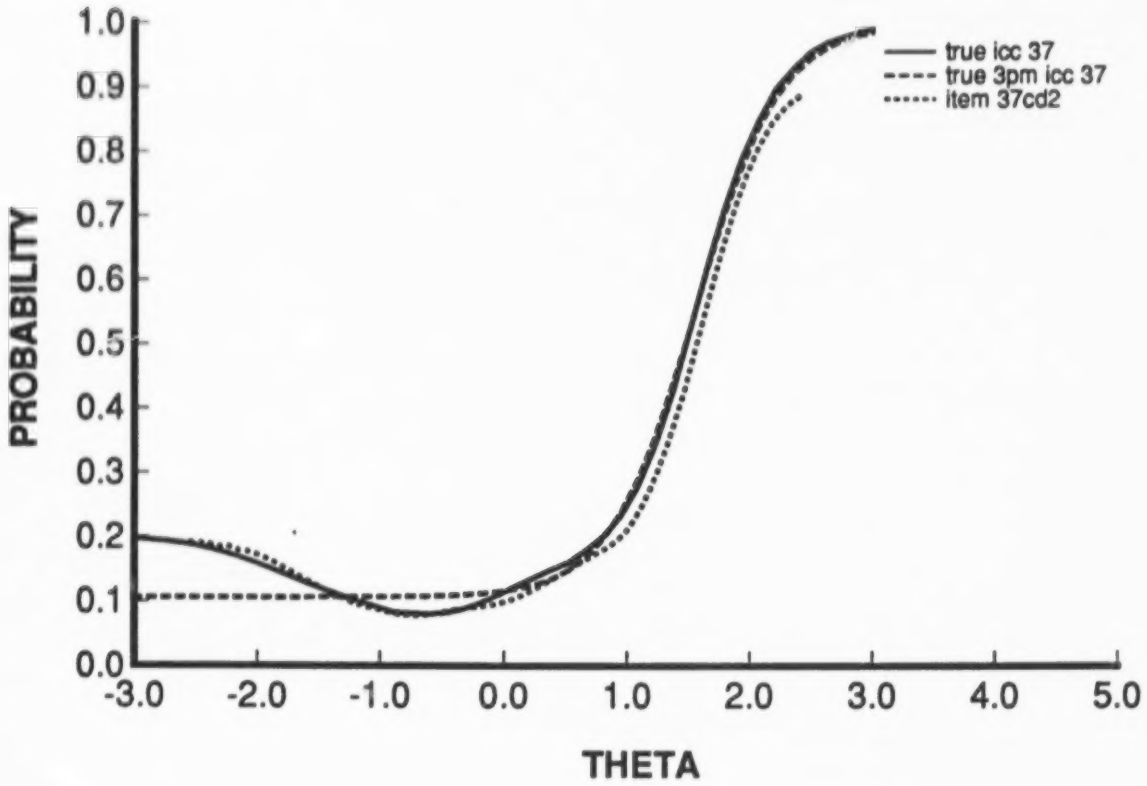
It can be shown that the test information function, $I(\theta)$, which is defined as the conditional expectation, given $\theta$, of the response pattern information function (Samejima, 1981), equals the sum total of the $n$ item information function, i.e.,

$$I(\theta) = \sum_{g=1}^{n} I_g(\theta)\ . \tag{5.3}$$

Let $\tau$ be a strictly increasing transformation of $\theta$, for which we can write

$$\tau = \tau(\theta) = C_1^{-1}\int_{-\infty}^{\theta}[I(t)]^{1/2}\,dt + C_0\ , \tag{5.4}$$

where $C_0$ is an arbitrary constant for adjusting the origin of $\tau$, and $C_1$ is an arbitrary positive value equal to the constant amount of test information for the range of $\tau$ of interst, respectively. In




0.641 0.80 1.50 6.00 6.51
CDR90037.DAT, INCDR9, plotted by NANCY DOMM

FIGURE 5-1

**Examples of the Estimated Operating Characteristic of the Correct Answer (Dotted Line) in Comparison with the True Operating Characteristic (Solid Line). Differential Weight Procedure of the Conditional P.D.F. Approach Based upon the Simple Sum Procedure Combined with the Normal Approach Method was Used for the Estimation. Also Presented is the Operating Characteristic in the Three-Parameter Logistic Model Fitted to the True Operating Characteristic by Dr. Michael Levine (Dashed Line).**

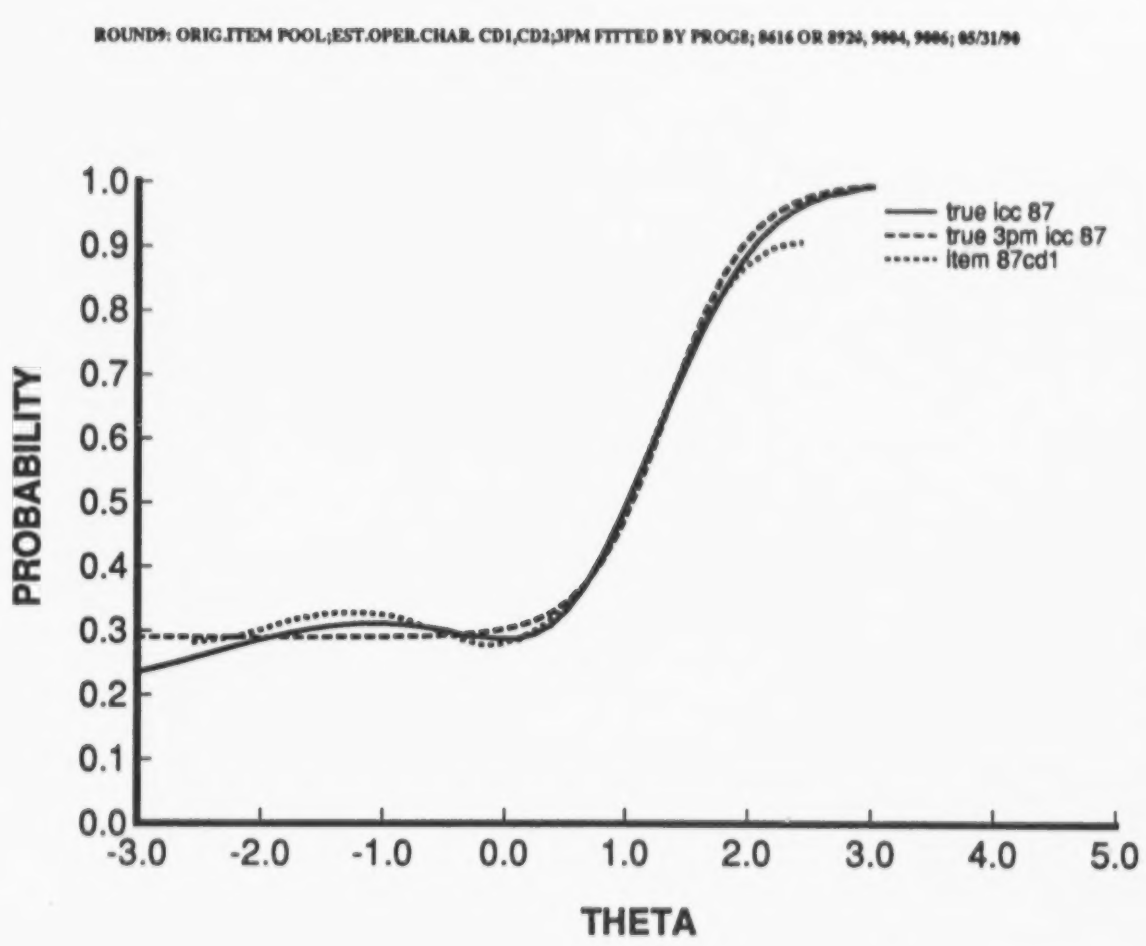



FIGURE 5-1 (Continued)

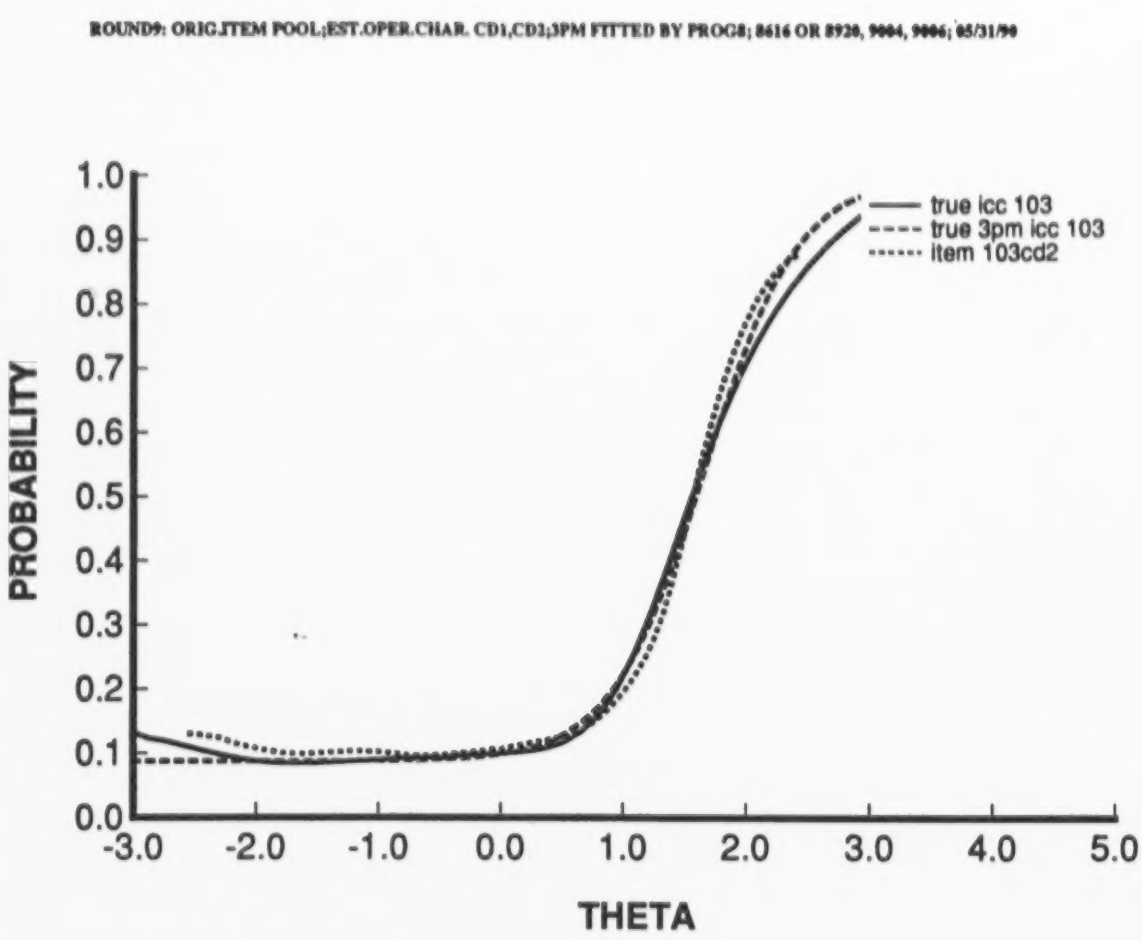
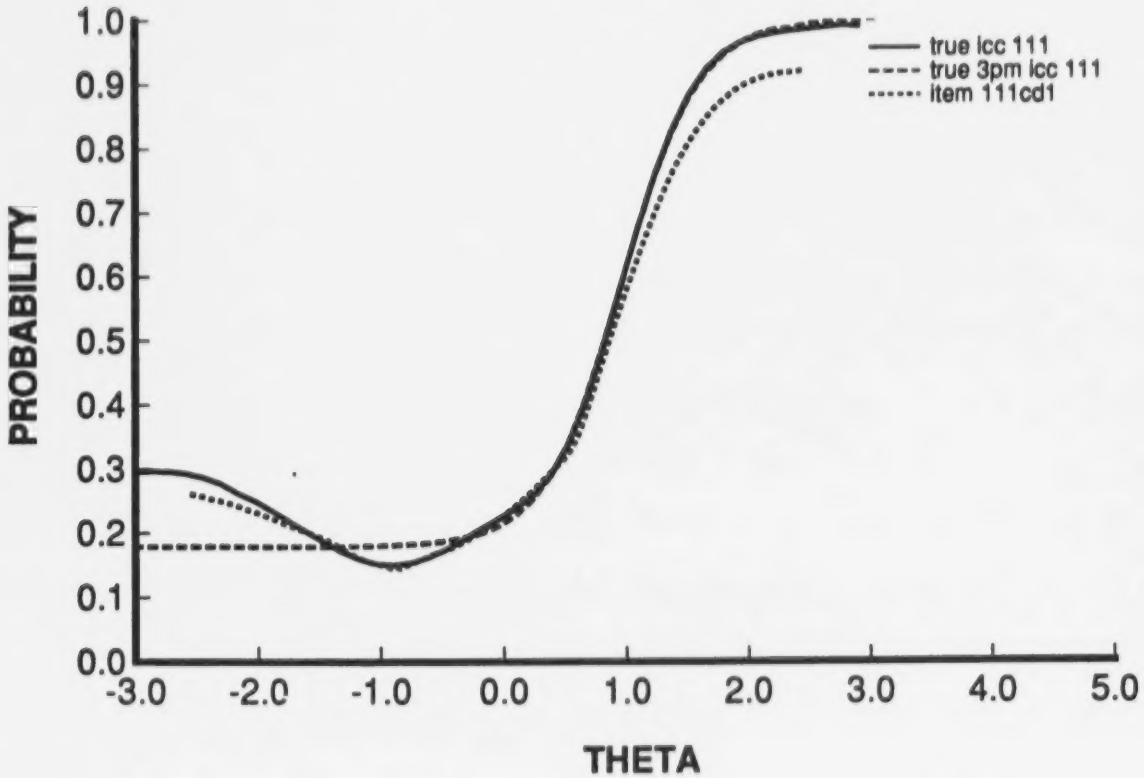



FIGURE 5-1 (Continued)

ROUND9: ORIG.ITEM POOL;EST.OPER.CHAR. CD1,CD2;3PM FITTED BY PROG8; 8616 OR 8920, 9004, 9006; 05/31/90




0.841 0.80 1.50 6.00 6.51
CDR90111.DAT, INCDR9, plotted by NANCY DOMM

FIGURE 5-1 (Continued)

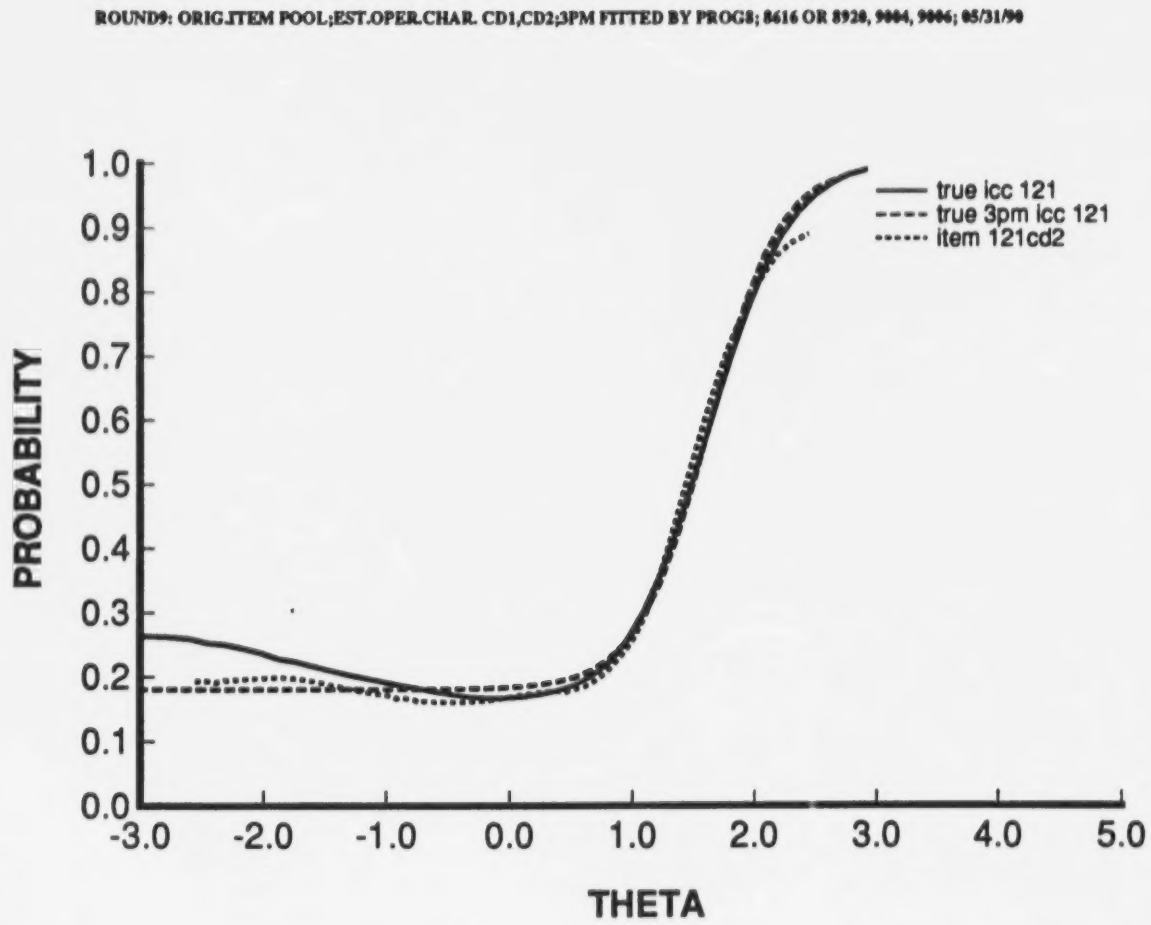



FIGURE 5-1 (Continued)

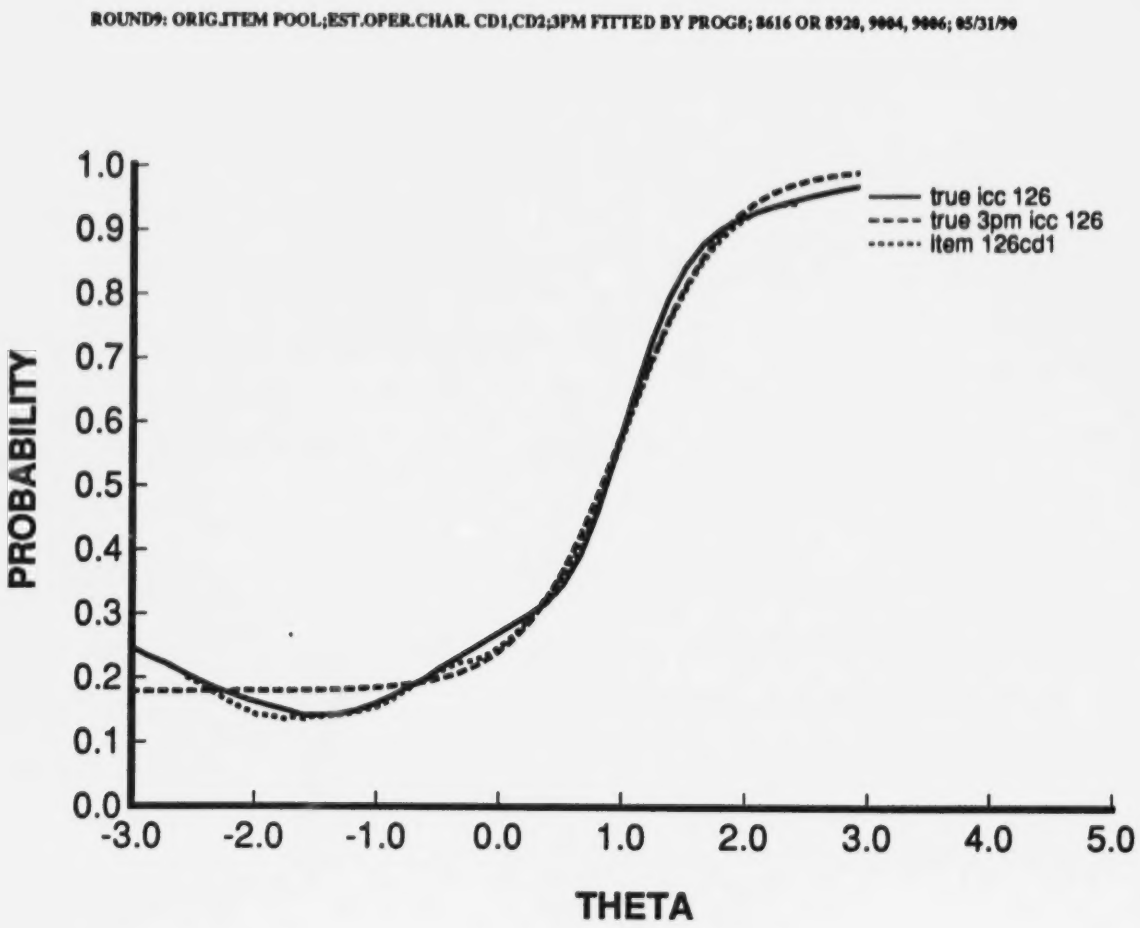
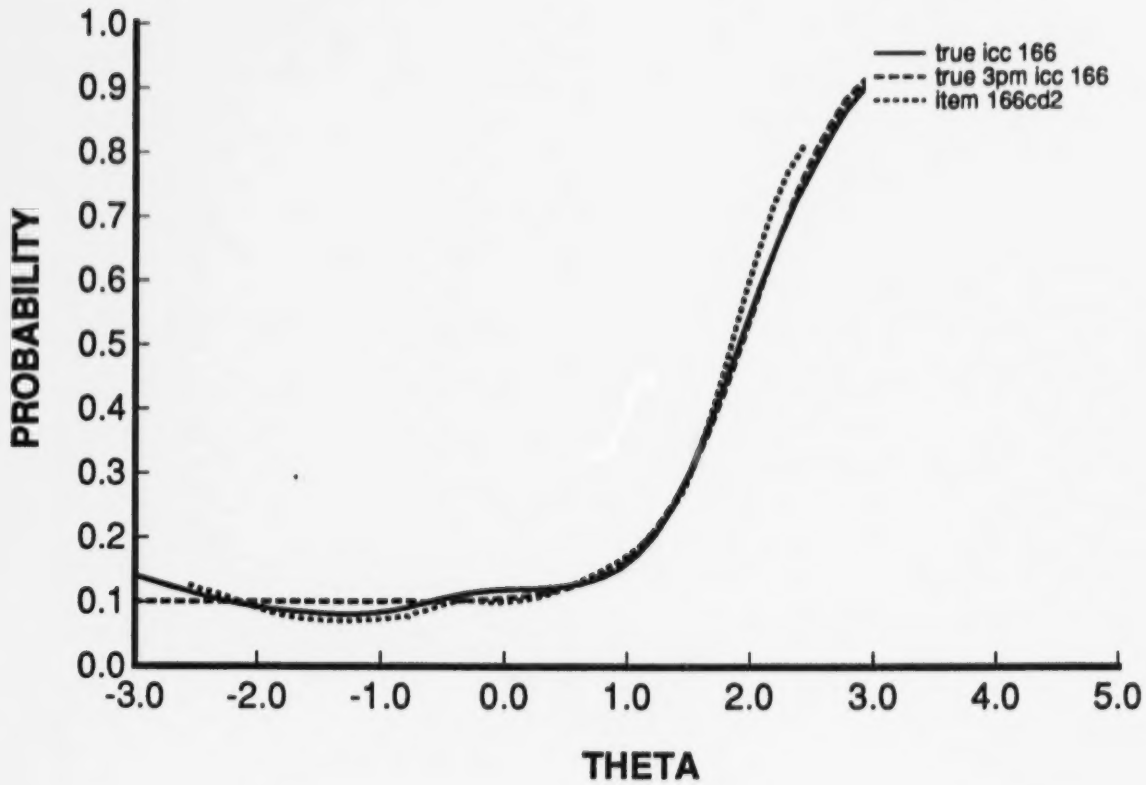
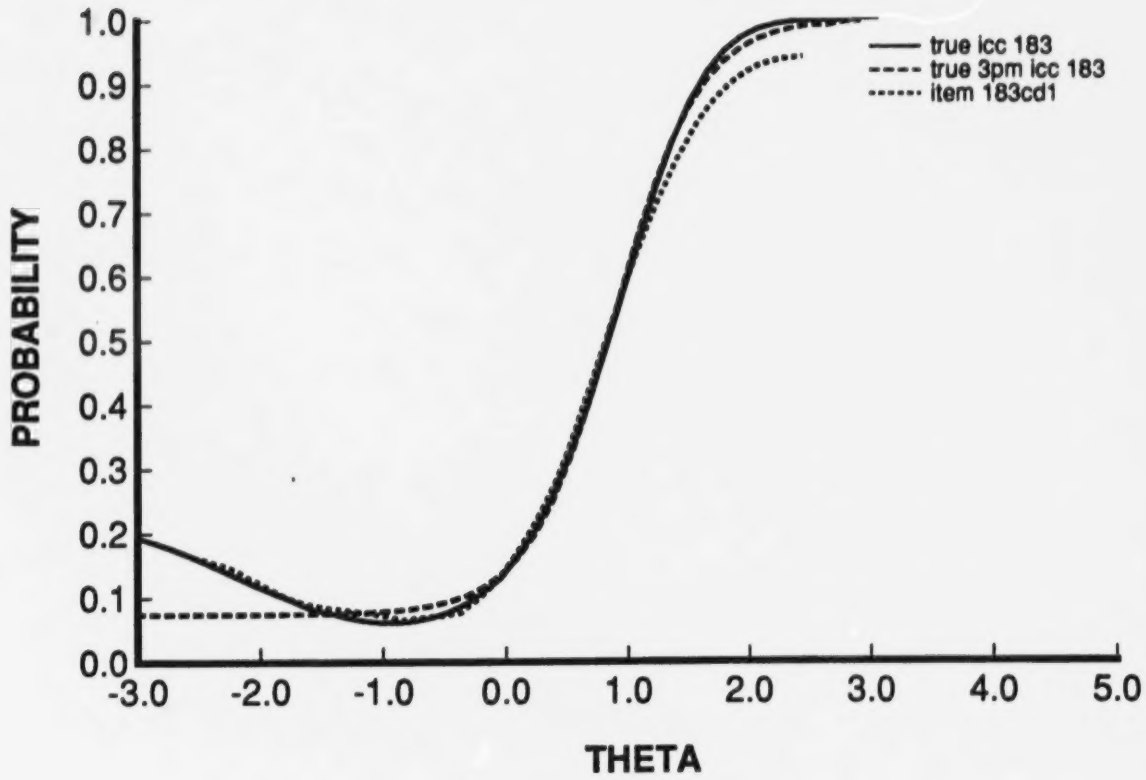



FIGURE 5-1 (Continued)




0.841 0.80 1.50 6.00 6.51
CDR90186.DAT, INCDR9, plotted by NANCY DOMM

FIGURE 5-1 (Continued)




0.841 0.80 1.50 6.00 6.51
CDR90183.DAT, INCDR9, plotted by NANCY DOMM

FIGURE 5-1 (Continued)

the Simple Sum Procedure of the Conditional P.D.F. Approach, the estimated operating characteristic $P^*_{k_h}(\theta)$ of the discrete response $k_h$ to an *unknown* item $h$ is obtained by

$$\hat{P}_{k_h}(\theta) = \hat{P}^*_{k_h}[\tau(\theta)] = \sum_{s \in k_h} \hat{\phi}(\tau \mid \hat{\tau}_s)[\sum_{s=1}^{N} \hat{\phi}(\tau \mid \hat{\tau}_s)]^{-1} \tag{5.5}$$

where $s$ is an individual, $N$ is the number of individuals, and $\hat{\phi}(\tau \mid \hat{\tau}_s)$ is the estimated conditional density of $\tau$, given the maximum likelihood estimate $\hat{\tau}_s$ of $\tau$ for the individual $s$. In the Normal Approach Method, $\hat{\phi}(\tau \mid \hat{\tau}_s)$ is obtained by fitting the normal density function to $\phi(\tau \mid \hat{\tau}_s)$, using the estimated first and second conditional moments of $\tau$, given $\hat{\tau}_s$, as its two parameters. We have

$$\hat{E}(\tau \mid \hat{\tau}) = \hat{\tau} + C_1^{-2} \frac{\partial}{\partial \hat{\tau}} \log\ \hat{g}(\hat{\tau}) \ , \tag{5.6}$$

and

$$\hat{V}ar.(\tau \mid \hat{\tau}) = C_1^{-2}\ [1 + C_1^{-2} \frac{\partial^2}{\partial \hat{\tau}^2} \log\ \hat{g}(\hat{\tau})] \ , \tag{5.7}$$

where $\hat{g}(\hat{\tau})$ is the estimated density function of $\hat{\tau}$, which can be approximated by a polynomial fitted by the method of moments (Samejima and Livingston, 1979).

In the Differential Weight Procedure of the Conditional P.D.F. Approach, we can write for the estimated operating characteristic

$$\hat{P}_{k_h}(\theta) = \hat{P}^*_{k_h}[\tau(\theta)] = \sum_{s \in k_h} W_{k_h}(\tau)\phi(\tau \mid \hat{\tau}_s)[\sum_{s=1}^{N} W_{k_h}(\tau; s)\phi(\tau \mid \hat{\tau}_s)]^{-1} \ , \tag{5.8}$$

where $W_{k_h}(\tau)$ is a differential weight function. Since this function involves $P^*_{k_h}(\tau)$ (cf. Samejima, 1990), which itself is the target of estimation, we may use its estimate obtained by the Simple Sum Procedure combined with the Normal Approach Method as its initial approximation, with some modifications.

Figure 5-1 presents eight examples of the estimated operating characteristic of the correct answer of the dichotomous test item obtained by using the Differential Weight Procedure. They are based upon simulated data provided by Dr. Charles Davis of the Office of Naval Research. The true operating characteristic of the correct answer for each item is available, therefore, and it is also drawn in each graph. In each graph, also presented is the operating characteristic in the three-parameter logistic model fitted to the true operating characteristic by Dr. Michael Levine. As you can see in this figure, Differential Weight Procedure of the Conditional P.D.F. Approach will provide us with fairly good estimates of operating characteristics, if we choose suitable differential weight functions.

An important implication of these results is that the nonparametric approach for estimating the operating characteristic has succeeded in approximating *non-monotonic* functions. This is essential in using any method for the estimation of plausibility functions. Although we need more research for improving the fitnesses further, these results give us promises for success in identifying informative distractors and in estimating their operating characteristics.

Item analysis has a long history, starting from the classical proportion correct and item-test regression. In the context of latent trait models, the operating characteristics and the information functions have provided us with powerful tools. Now we can add the plausibility functions of the distractors to this category. By accurately identifying the configuration of the operating characteristics of the correct answer and the distractors, we shall be able to understand the characteristics of the item, its strengths

and weaknesses. In this way modifications of the item can be done if necessary. Successful nonparametric methods of estimating the operating characteristics are essential, therefore, for this new, more informative approach to the item analysis.

# VI Efficiency in Ability Estimation and Strategies of Writing Test Items

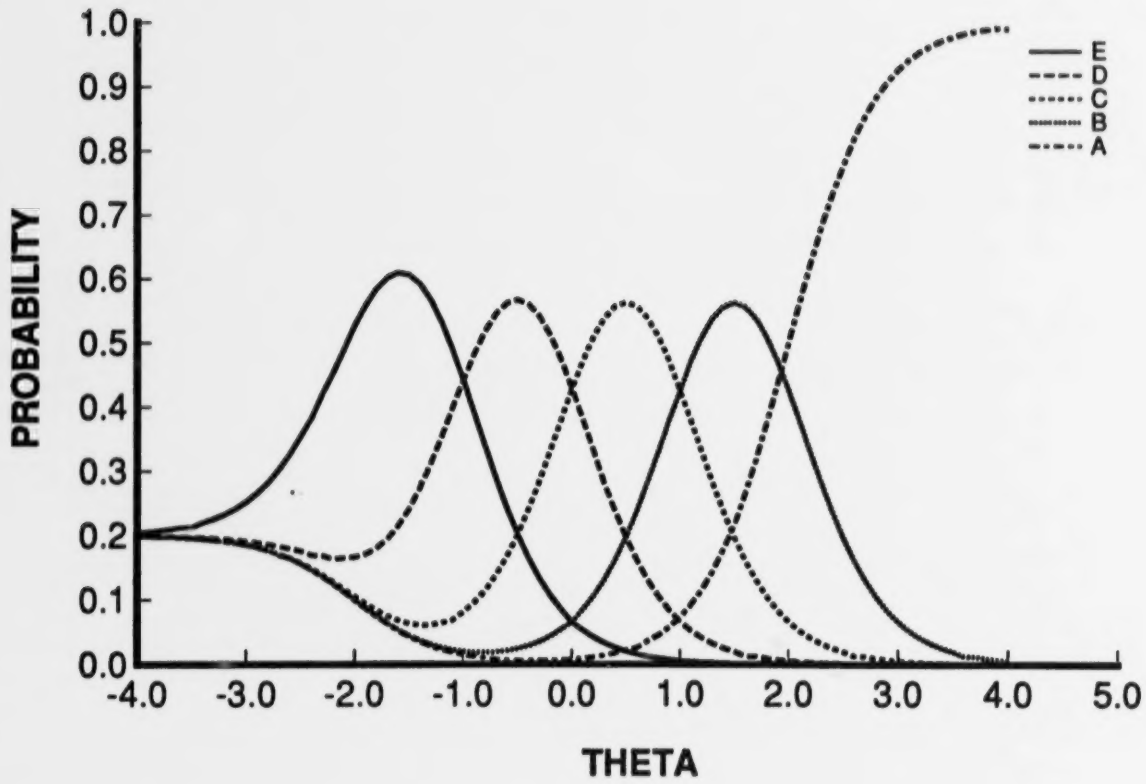
Observations and discussion made in the preceding sections give us many useful informations as well as warnings. First of all, theoretical observations indicate that non-monotonicity of the operating characteristic of the correct answer to the multiple-choice test item is a natural consequence of theory. Secondly, it has been shown from several different angles that the third parameter, $c_g$, in the three-parameter model provides us with nothing but noise; the greater the value of $c_g$ the more noise and inaccuracies in estimation it produces. Thirdly, it has been pointed out that, although it is still a common procedure for researchers to *mold* the operating characteristics of the correct answers of their multiple-choice test items into the three-parameter logistic model, some nonparametric methods applied to *empirical* data have revealed the non-monotonicity of the operating characteristic of the correct answer with many *actual* test items, as well as differential informations provided by separate distractors. Fourthly, it has been pointed out that nonparametric approach to the estimation of the operating characteristics of discrete responses has been successful enough to detect the non-monotonicity of the function when it exists, and to approximate their rather irregular curves fairly accurately.

With all these facts, it is time to reconsider conventional strategies for item writing and to propose new strategies.

The first thing we need to reconsider is the lack of sufficient interactions between theorists and people who write test items. It has been fairly common that: 1) a committee is organised for writing test items in a specified content area or domain and eventually produces a set of test items; 2) another group of people tests these items on a small sample of subjects, screens the items and then administers the selected items to larger groups of subjects. Item calibration is done on the second stage, assuming some model such as the three-parameter logistic model, etc. In most cases, there is practically no feedback from theorists to item writers. If we set a strategy that more interactions are made between the two groups of people so that the test items are revised and pilot tested with each interaction, we shall be able to improve the test, and the improvement will lead to efficiency in ability estimation.

The second thing we need to reconsider is the simpleminded avoidance of non-monotonicity of the operating characteristic of the correct answer. While it is not desirable for an item to have higher conditional probabilities of the correct answer on lower levels of ability than on higher levels, selecting alternative answers so that the "dips" of the operating characteristic of the correct answer be *smoothed out* will lead to a substantially large value of the lower asymptote of the operating characteristic in most cases. We must recall that even a small number like 0.2 as $c_g$ in the three-parameter logistic model is a big nuisance, as was discussed in Section 3. Our strategy must be that we make the best use of those "dips", instead of avoiding them.

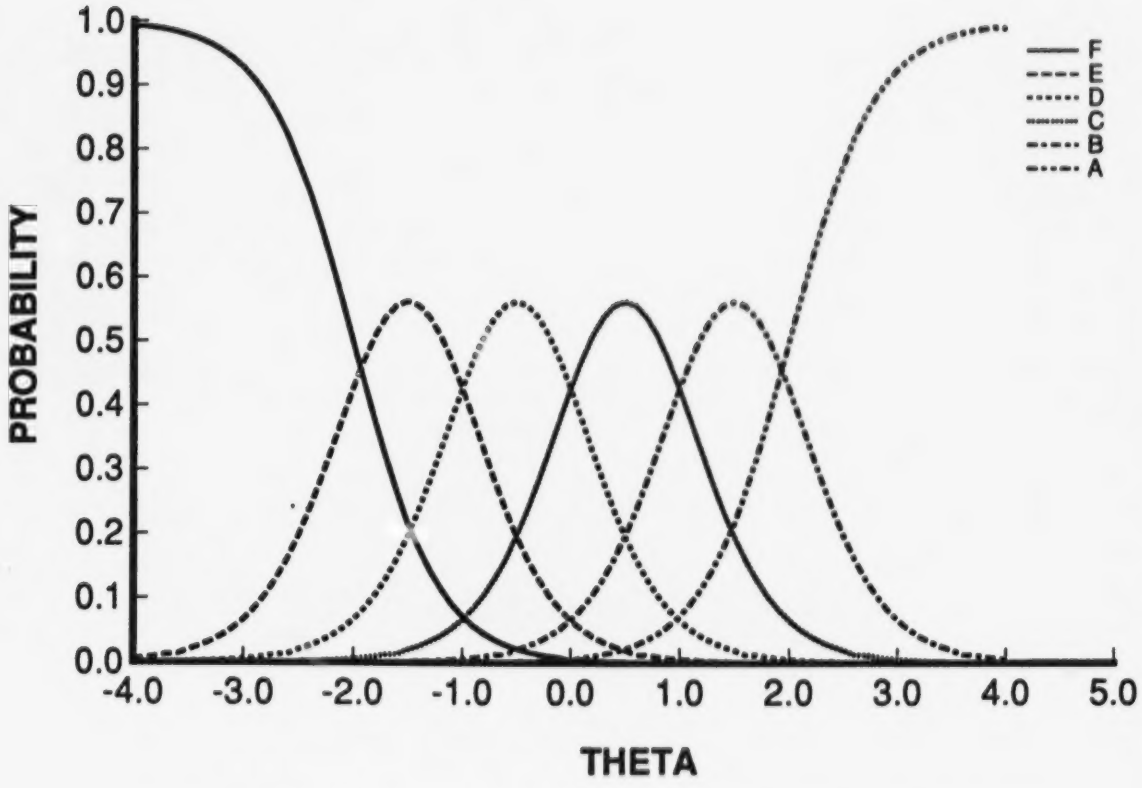
Figure 6-1 presents the operating characteristics of the five alternative answers of a hypothesised test item following Model B (Samejima, 1979b), with the parameter values: $a_g = 1.50$, $b_1 = -2.00$, $b_2 = -1.00$, $b_3 = 0.00$, $b_4 = 1.00$ and $b_5 = 2.00$. The subscript for each of the five difficulty parameters indicates the order of *easiness* for the examinee to be attracted to the plausibility of each alternative answer, so that, in this example, $b_5$ indicates the difficulty parameter of the correct answer. We can see in this figure that a practical monotonicity exists for the operating characteristic of the correct answer for the range of $\theta$, $(-0.5, \infty)$, and, more importantly, within this range of $\theta$ its lower asymptote is very close to zero, i.e., the nuisance caused by the non-zero lower asymptote will be gone as far as we administer the item to populations of subjects whose ability distributes on higher




0.641 0.80 1.50 6.00 6.51
7933IT1.DAT, IN7933, plotted by NANCY DOMM

FIGURE 6-1

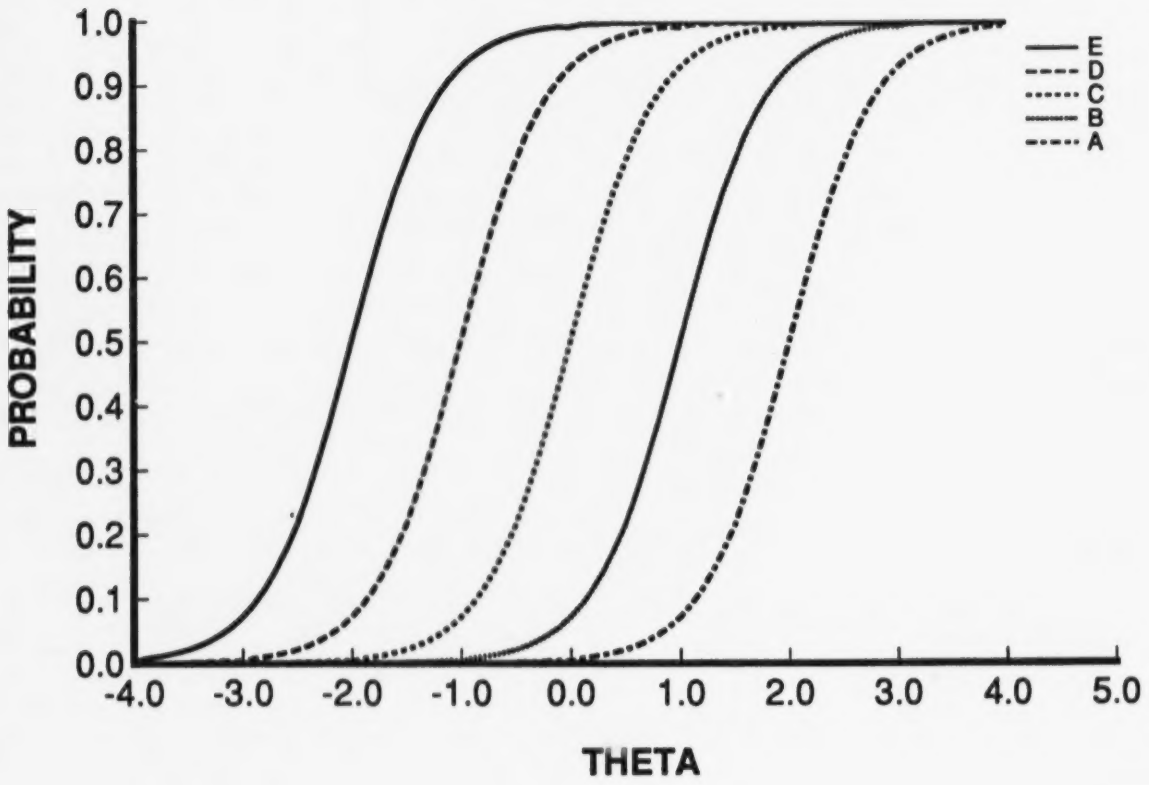
Operating Characteristics of the Five Alternative Answers of a Hypothetical Test Item Following Model B, with the Parameter Values: $a_g = 1.5$ , $b_1 = -2.0$ , $b_2 = -1.0$ , $b_3 = 0.0$ , $b_4 = 1.0$ and $b_5 = 2.0$ .




0.841 0.80 1.50 8.00 6.51
7933IT1.DAT, IN7933, plotted by NANCY DOMM

FIGURE 6-2

Operating Characteristics of the Five Alternative Answers of a Hypothetical Test Item in the Free-Response Situation Following the Logistic Model on the Graded Response Level, with the Parameter Values: $a_g = 1.5$ , $b_1 = -2.0$ , $b_2 = -1.0$ , $b_3 = 0.0$ , $b_4 = 1.0$ and $b_5 = 2.0$ .




0.641 0.80 1.50 8.00 8.51
7933IT1.DAT, IN7933, plotted by NANCY DOMM

FIGURE 6-3

**Operating Characteristics of the Correct Answer Obtained by the Five Different Redichotomisations of the Graded Test Item Following the Logistic Model, with the Discrimination Parameter, $a_g = 1.5$, and the Difficulty Parameters, $b_1 = -2.0$, $b_2 = -1.0$, $b_3 = 0.0$, $b_4 = 1.0$ and $b_5 = 2.0$, Respectively.**

levels than $\theta = -0.5$ .

These operating characteristics of the five alternative answers in Figure 6-1 are originated from those in the logistic model on the graded response level (Samejima, 1969, 1972) with the same parameter values (cf. Samejima, 1979b). Figures 6-2 presents the corresponding set of operating characteristics of the correct answers in the logistic model. We notice there is an additional strictly decreasing curve in this figure. This curve represents the conditional probability, given $\theta$ , that the examinee does not find *attractiveness* in any alternative answers. In Model B, these people are assumed to guess randomly, so in Figure 6-1 this curve does not exist, and the conditional probability is evenly distributed among the five alternative answers to account for the rises in their operating characteristics at lower levels of $\theta$ .

Figure 6-3 presents the operating characteristics of the correct answer following the logistic model on the dichotomous response level, which are obtained by the five different redichotomisations of the graded test item exemplified in Figure 6-2. In these functions, $a_g = 1.5$ is the common discrimination parameter, and the difficulty parameters are: $b_g = -2.0, -1.0, 0.0, 1.0, 2.0$ , respectively. This is the starting point of the graded response model, which leads to the operating characteristics illustrated in Figure 6-2 (cf. Samejima, 1969, 1972).

Suppose that two alternative answers which attract examinees of low levels of $\theta$ are replaced, and the revised item has $b_1 = -3.00$ and $b_2 = -1.50$ , respectively. In this situation, the operating characteristics of the correct answer obtained by the first two redichotomisations are changed, as are shown in Figure 6-4. This revision leads to the set of operating characteristics in the logistic model on the graded response level presented by Figure 6-5. If we compare this figure with Figure 6-2, we can see that the curve for the category of *not attracted* examinees is shifted to substantially lower levels of $\theta$ . Figure 6-6 presents the set of operating characteristics for this revised test item following Model B. In this figure we can see that the operating characteristic of the correct answer is practically strictly increasing within the range of $\theta$ , $(-1.7, \infty)$ , and the *pseudo* lower asymptote of the operating characteristic within this range of $\theta$ is still very close to zero.

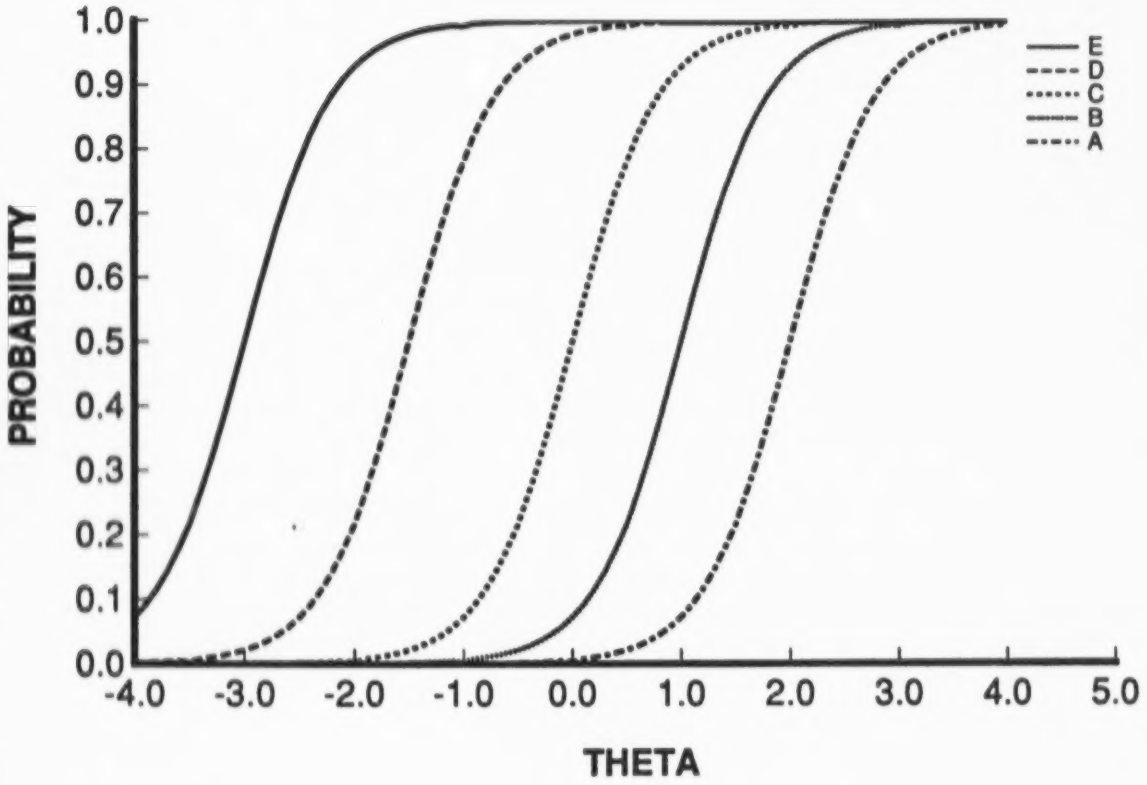
A big gain resulting from this revision is the fact that the lower endpoint of the interval of $\theta$ in which the operating characteristic of the correct answer is practically monotonic has substantially shifted to the negative direction, while still keeping its lower asymptote practically zero. Thus we can avoid the noise coming from the lower asymptote even if we administer the item to populations of examinees whose ability distributions are located on lower levels of $\theta$ . In other words, without sacrificing the accuracy of ability estimation, the utility of the item has been substantially enhanced by this revision.

The above example suggests the following strategy.

(1) If the nonparametrically estimated operating characteristic of the correct answer to an item provides us with a relatively high value of $\theta$ below which monotonicity does not exist, then change the set of distractors to include one or more wrong answers that attract examinees of very low levels of ability.

It may sound difficult to do in practice. If we pay attention to actually used multiple-choice test items, however, we will come across many wrong alternative answers that are attracting examinees of very low levels of ability. To give an example, the author has come across an arithmetic item asking for the area of a rectangle. A substantial number of seventh graders chose the wrong alternative answer which equals the sum of the two sides of the rectangle of different lengths! It is obvious that those who did not understand how to obtain the area of a rectangle at all chose this alternative answer.

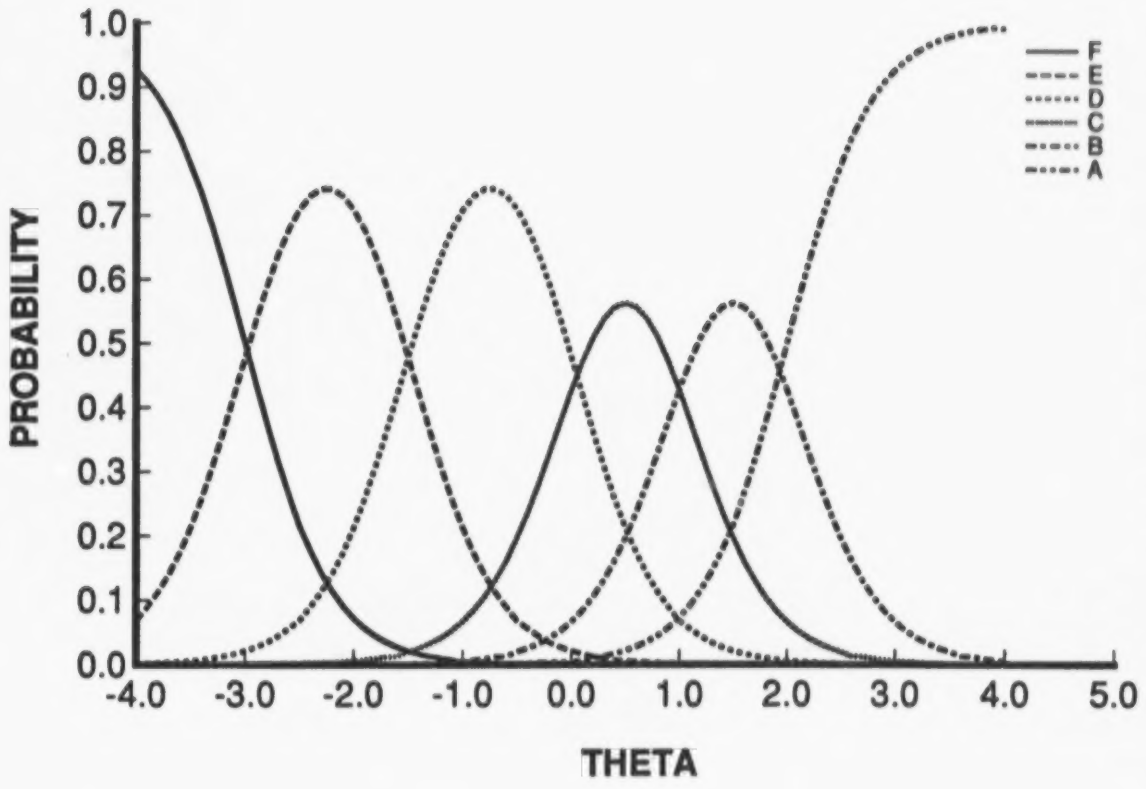
Another consideration which is important in writing test items is to keep the "pseudo" lower asymptote of the operating characteristic of the correct answer *close enough to zero*, as is the case with the above example. This has a great deal to do with the discrimination powers of the alternative answers, as well as the configuration of the plausibility functions. Figures 6-7 through 6-9 present, in the reversed order, the same set of three figures as Figures 6-1 through 6-3, by changing the discrimination




0.841 0.80 1.50 6.00 6.51
7933IT2.DAT, IN7933, plotted by NANCY DOMM

FIGURE 6-4

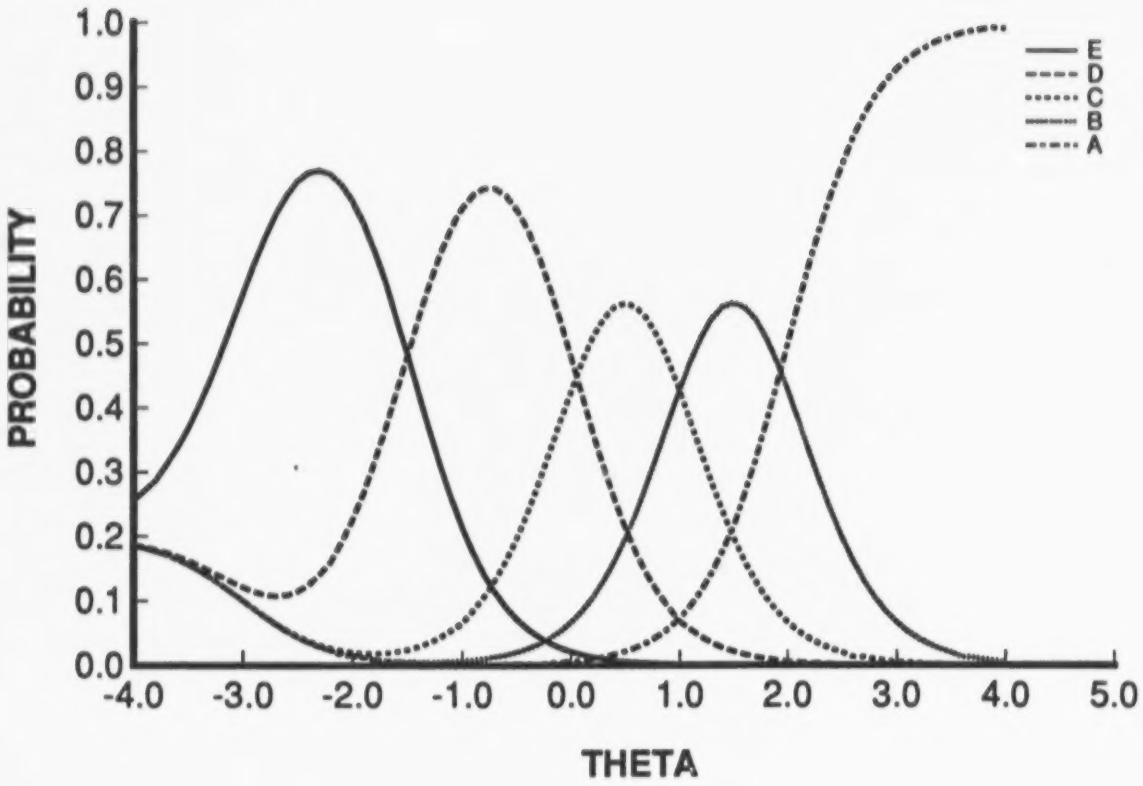
**Operating Characteristics of the Correct Answer Obtained by the Five Different Redichotomisations of the Graded Test Item Following the Logistic Model, with the Discrimination Parameter, $a_g = 1.5$, and the Difficulty Parameters, $b_1 = -3.0$, $b_2 = -1.5$, $b_3 = 0.0$, $b_4 = 1.0$ and $b_5 = 2.0$, Respectively.**




0.641 0.80 1.50 6.00 6.51
7933IT2.DAT, IN7933, plotted by NANCY DOMM

FIGURE 6-5

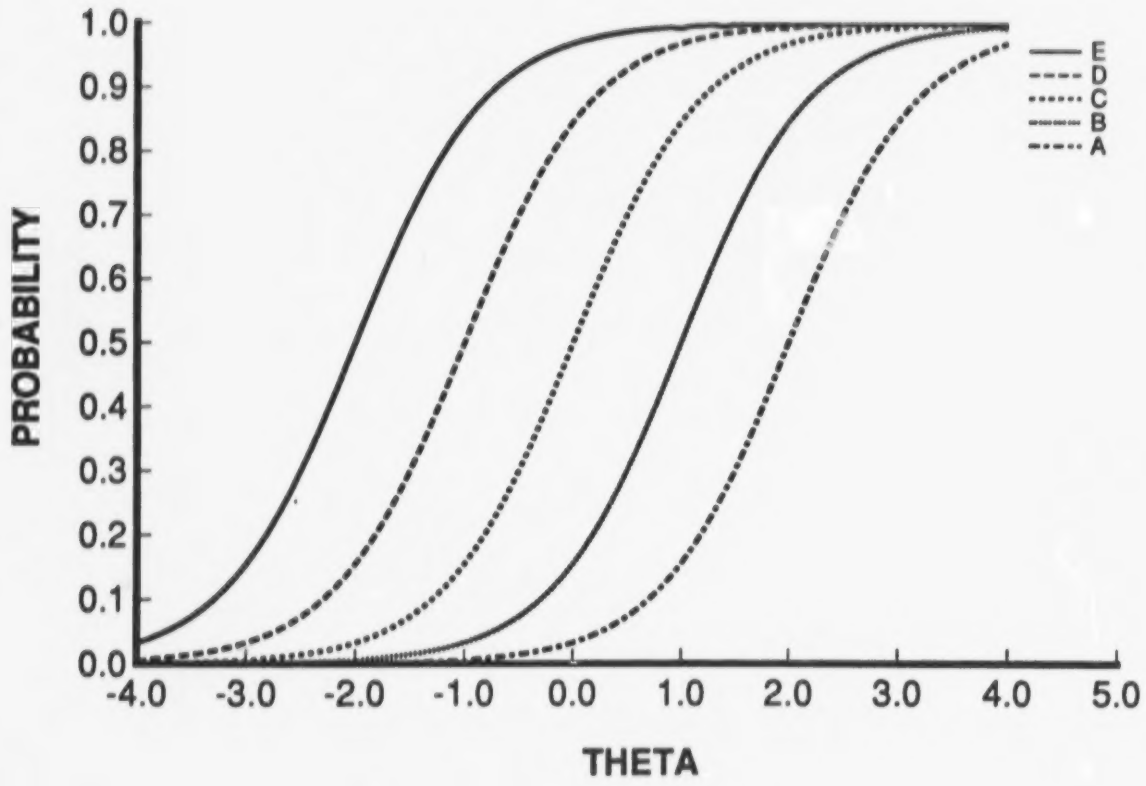
**Operating Characteristics of the Five Alternative Answers of a Hypothetical Test Item in the Free-Response Situation Following the Logistic Model on the Graded Response Level, with the Parameter Values: $a_g = 1.5$, $b_1 = -3.0$, $b_2 = -1.5$, $b_3 = 0.0$, $b_4 = 1.0$ and $b_5 = 2.0$.**




0.841 0.80 1.50 6.00 6.51
7933IT2.DAT, IN7933, plotted by NANCY DOMM

FIGURE 6-6

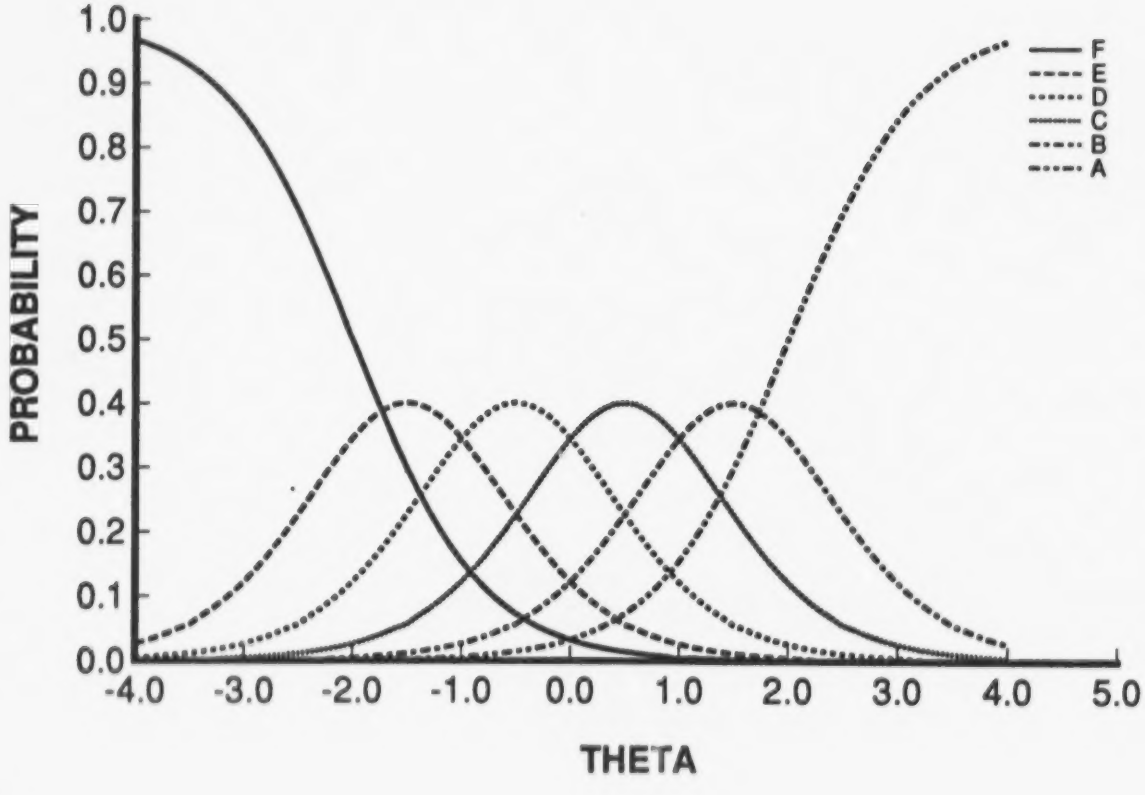
**Operating Characteristics of the Five Alternative Answers of a Hypothetical Test Item Following Model B, with the Parameter Values: $a_g = 1.5$ , $b_1 = -3.0$ , $b_2 = -1.5$ , $b_3 = 0.0$ , $b_4 = 1.0$ and $b_5 = 2.0$ .**




0.641 0.80 1.60 6.00 6.51
7933IT3.DAT, IN7933, plotted by NANCY DOMM

FIGURE 6-7

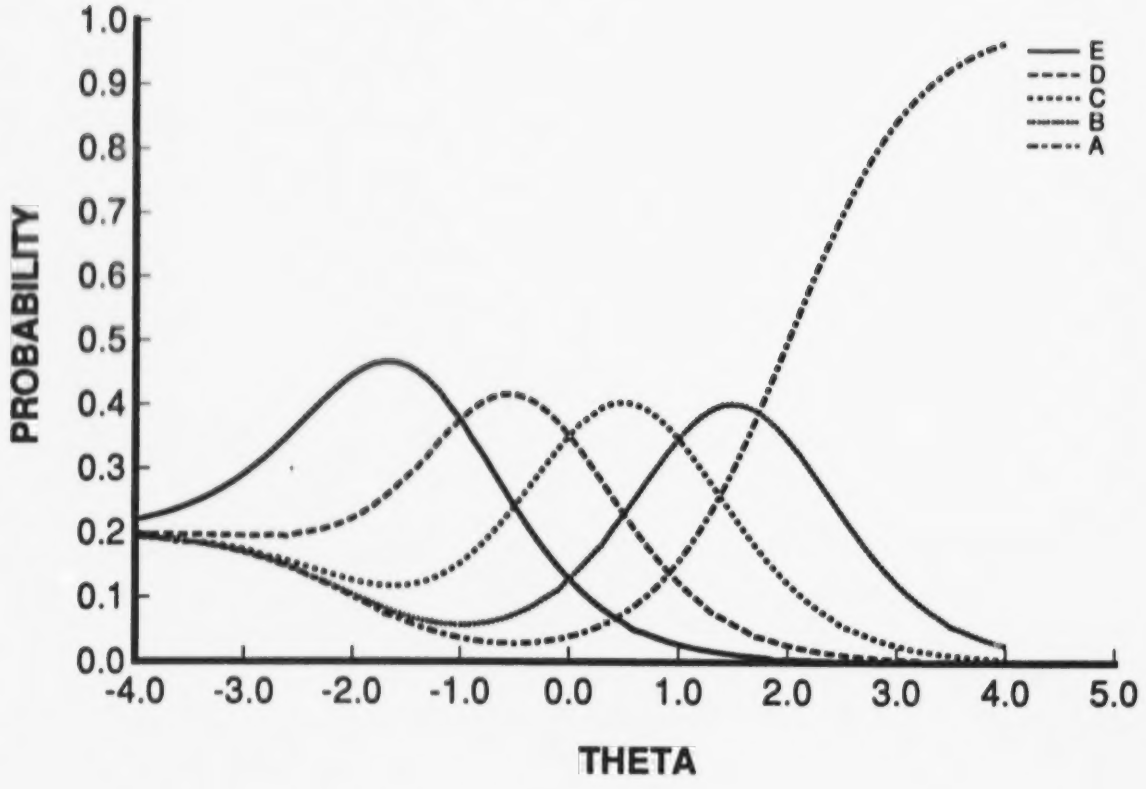
Operating Characteristics of the Correct Answer Obtained by the Five Different Redichotomisations of the Graded Test Item Following the Logistic Model, with the Discrimination Parameter, $a_g = 1.0$, and the Difficulty Parameters, $b_1 = -2.0$, $b_2 = -1.0$, $b_3 = 0.0$, $b_4 = 1.0$ and $b_5 = 2.0$, Respectively.




0.641 0.80 1.50 6.00 6.51
7933IT3.DAT, IN7933, plotted by NANCY DOMM

FIGURE 6-8

Operating Characteristics of the Five Alternative Answers of a Hypothetical Test Item in the Free-Response Situation Following the Logistic Model on the Graded Response Level, with the Parameter Values: $a_g = 1.0$, $b_1 = -2.0$, $b_2 = -1.0$, $b_3 = 0.0$, $b_4 = 1.0$ and $b_5 = 2.0$.




0.641 0.80 1.50 6.00 6.51
7933IT3.DAT, IN7933, plotted by NANCY DOMM

FIGURE 6-9

Operating Characteristics of the Five Alternative Answers of a Hypothetical Test Item Following Model B, with the Parameter Values: $a_g = 1.0$, $b_1 = -2.0$, $b_2 = -1.0$, $b_3 = 0.0$, $b_4 = 1.0$ and $b_5 = 2.0$.

parameter from $a_g = 1.5$ to $a_g = 1.0$, while keeping the five difficulty parameters unchanged. If we compare Figure 6-9 with Figure 6-1, we can see a substantial enhancement of the "pseudo" lower asymptote within the interval of $\theta$, $(-0.5, \infty)$, i.e., the nuisance has been increased by the change of the discrimination parameter.

This suggests the second strategy:

(2) If possible, try to include distractors whose estimated operating characteristics are *steep*, while keeping the differential configuration of the these functions, as is illustrated in Figure 6-4.

So far our strategies have been focused upon producing an informative operating characteristic of the correct answer. We notice, however, that *these strategies will also provide us with distractors which provide us with differential informations*. This implies that approximation of the nonparametrically estimated operating characteristics of one or more alternative answers by some mathematical formulae will enable us to use these additional differential informations in ability estimation. This *posterior parameterization* of the non-parametrically estimated operating characteristics of distractors will certainly lead us to the increased accuracy and efficiency in ability measurement.

## VII Discussion and Conclusions

The present paper summarises the shortages of the conventional way of handling the multiple-choice test and also describes theories and methodologies that can be applied for a better handling of the multiple-choice test item; some empirical facts are introduced to support the theoretical observations; finally, new strategies of item writing are proposed which will reduce *noise* and lead to more efficient ability estimation.

In spite of many controversies against the multiple-choice test, because of its economy in scoring it has been, and still is, very popular among people of psychological and educational measurement. Fortunately, theorists in mathematical psychology have developed many new ideas and methodologies in the past couple of decades that can improve the way of handling the multiple-choice test. Nonparametric approach in estimating the operating characteristic is one of them. Also the rapid progress in electronic technologies has made it possible to materialise these results of theories and methodologies in practical situations. Today, we are in a position to take advantage of all these accomplishments.

## References


[1] Bock, R. D. and Aitkin, M. Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika, 46*, 1981, 443-459.

[2] Lord, F. M. *Applications of item response theory to practical testing problems*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1980.

[3] Lord, F. M. and Novick, M. R. *Statistical theories of mental test scores*. Reading, Mass.: Addison-Wesley, 1968.

[4] Samejima, F. Application of the graded response model to the nominal response and multiple-choice situations. *UNC Psychometric Laboratory Report, 63*, 1968.

[5] Samejima, F. Estimation of ability using a response pattern of graded scores. *Psychometrika Monograph, No. 17*, 1969.

[6] Samejima, F. A general model for free-response data. *Psychometrika Monograph, No. 18*, 1972.

[7] Samejima, F. A comment on Birnbaum's three-parameter logistic model in the latent trait theory. *Psychometrika, 38*, 1973, 221-233.

[8] Samejima, F. Constant information model: A new, promising item characteristic function. *ONR/RR-79-1*, 1979a.

[9] Samejima, F. A new family of models for the multiple-choice item. *ONR/RR-79-4*, 1979b.

[10] Samejima, F. Final Report: Efficient methods of estimating the operating characteristic of item response categories and challenge to a new model for the multiple-choice item. *Final Report of N00014-77-C-0360*, Office of Naval Research, 1981.

[11] Samejima, F. Information loss caused by noise in models for dichotomous items. *ONR/RR-82-1*, 1982a.

[12] Samejima, F. Effect of noise in the three-parameter logistic model. *ONR/RR-82-2*, 1982b.

[13] Samejima, F. Plausibility functions of Iowa Vocabulary Test items estimated by the Simple Sum Procedure of the Conditional P.D.F. Approach. *ONR/RR-84-1*, 1984a.

[14] Samejima, F. Results of item parameter estimation using Logist 5 on simulated data. *ONR/RR-84-3*, 1984b.

[15] Samejima, F. Differential Weight Procedure of the Conditional P.D.F. Approach for estimating the operating characteristics of discrete item responses. *ONR/RR-90-4*, 1990.

[16] Samejima, F. and Livingston, P. Method of moments as the least squares solution for fitting a polynomial. *ONR/RR-79-2*, 1979.

[17] Wingersky, M. S., Barton, M. A., and Lord, F. M. *Logist user's guide*. Princeton: Educational Testing Service, 1982.

[18] Yen, W. M., Burket, G. R., and Sykes, R. C. Nonunique solutions to the likelihood equation for the three-parameter logistic model. *Psychometrika*, in press.

ONR90051.TEX;ONR90052.TEX
June 5, 1990

Distribution List

Dr. Terry Ackerman
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. James Algina
1403 Norman Hall
University of Florida
Gainesville, FL 32605

Dr. Erling B. Andersen
Department of Statistics
Studiestraede 6
1455 Copenhagen
DENMARK

Dr. Ronald Armstrong
Rutgers University
Graduate School of Management
Newark, NJ 07102

Dr. Eva L. Baker
UCLA Center for the Study
of Evaluation
145 Moore Hall
University of California
Los Angeles, CA 90024

Dr. Laura L. Barnes
College of Education
University of Toledo
2801 W. Bancroft Street
Toledo, OH 43606

Dr. William H. Bart
University of Minnesota
Dept. of Educ. Psychology
330 Burton Hall
178 Pillsbury Dr., S.E.
Minneapolis, MN 55455

Dr. Isaac Bejar
Mail Stop: 10-R
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Dr. Menucha Birenbaum
School of Education
Tel Aviv University
Ramat Aviv 69978
ISRAEL

Dr. Arthur S. Blaives
Code N712
Naval Training Systems Center
Orlando, FL 32813-7100

Dr. Bruce Bloxom
Defense Manpower Data Center
99 Pacific St.
Suite 155A
Monterey, CA 93943-3231

Cdt. Arnold Bohrer
Sectie Psychologisch Onderzoek
Rekruterings-En Selectiecentrum
Kwartier Koningen Astrid
Bruijnstraat
1120 Brussels, BELGIUM

Dr. Robert Breaux
Code 281
Naval Training Systems Center
Orlando, FL 32826-3224

Dr. Robert Brennan
American College Testing
Programs
P. O. Box 168
Iowa City, IA 52243

Dr. John B. Carroll
409 Elliott Rd., North
Chapel Hill, NC 27514

Dr. John M. Carroll
IBM Watson Research Center
User Interface Institute
P.O. Box 704
Yorktown Heights, NY 10598

Dr. Robert M. Carroll
Chief of Naval Operations
OP-01B2
Washington, DC 20350

Dr. Raymond E. Christal
UES LAMP Science Advisor
AFHRL/MOEL
Brooks AFB, TX 78235

Mr. Hua Hua Chung
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Dr. Norman Cliff
Department of Psychology
Univ. of So. California
Los Angeles, CA 90089-1061

Director, Manpower Program
Center for Naval Analyses
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Director,
Manpower Support and
Readiness Program
Center for Naval Analysis
2000 North Beauregard Street
Alexandria, VA 22311

Dr. Stanley Collyer
Office of Naval Technology
Code 222
800 N. Quincy Street
Arlington, VA 22217-5000

Dr. Hans F. Crombag
Faculty of Law
University of Limburg
P.O. Box 616
Maastricht
The NETHERLANDS 6200 MD

Ms. Carolyn R. Crone
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

Dr. Timothy Davey
American College Testing Program
P.O. Box 168
Iowa City, IA 52243

Dr. C. M. Dayton
Department of Measurement
Statistics & Evaluation
College of Education
University of Maryland
College Park, MD 20742

Dr. Ralph J. DeAyala
Measurement, Statistics,
and Evaluation
Benjamin Bldg., Rm. 4112
University of Maryland
College Park, MD 20742

Dr. Lou DiBello
CERL
University of Illinois
103 South Mathews Avenue
Urbana, IL 61801

Dr. Dattprasad Divgi
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. Hei-Ki Dong
Bell Communications Research
6 Corporate Place
PYA-1K226
Piscataway, NJ 08854

Dr. Fritz Drasgow
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Defense Technical
Information Center
Cameron Station, Bldg 5
Alexandria, VA 22314
(12 Copies)

5/1/90

Dr. Stephen Dunbar
224B Lindquist Center
for Measurement
University of Iowa
Iowa City, IA 52242

Dr. James A. Earles
Air Force Human Resources Lab
Brooks AFB, TX 78235

Dr. Susan Embretson
University of Kansas
Psychology Department
426 Fraser
Lawrence, KS 66045

Dr. George Englehard, Jr.
Division of Educational Studies
Emory University
210 Fishburne Bldg.
Atlanta, GA 30322

ERIC Facility-Acquisitions
2440 Research Blvd, Suite 550
Rockville, MD 20850-3238

Dr. Benjamin A. Fairbank
Operational Technologies Corp.
5825 Callaghan, Suite 225
San Antonio, TX 78228

Dr. Marshall J. Farr, Consultant
Cognitive & Instructional
Sciences
2520 North Vernon Street
Arlington, VA 22207

Dr. P-A. Federico
Code 51
NPRDC
San Diego, CA 92152-6800

Dr. Leonard Feldt
Lindquist Center
for Mensurement
University of Iowa
Iowa City, IA 52242

Dr. Richard L. Ferguson
American College Testing
P.O. Box 168
Iowa City, IA 52243

5/1/90



Dr. Gerhard Fischer
Liebiggasse 5/3
A 1010 Vienna
AUSTRIA

Dr. Myron Fischl
U.S. Army Headquarters
DAPE-MRR
The Pentagon
Washington, DC 20310-0300

Prof. Donald Fitzgerald
University of New England
Department of Psychology
Armidale, New South Wales 2351
AUSTRALIA

Mr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Alfred R. Fregly
AFOSR/NL, Bldg. 410
Bolling AFB, DC 20332-6448

Dr. Robert D. Gibbons
Illinois State Psychiatric Inst.
Rm 529W
1601 W. Taylor Street
Chicago, IL 60612

Dr. Janice Gifford
University of Massachusetts
School of Education
Amherst, MA 01003

Dr. Drew Gitomer
Educational Testing Service
Princeton, NJ 08541

Dr. Robert Glaser
Learning Research
& Development Center
University of Pittsburgh
3939 O'Hara Street
Pittsburgh, PA 15260

Dr. Sherrie Gott
AFHRL/MOMJ
Brooks AFB, TX 78235-5601

Dr. Bert Green
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

Michael Habon
DORNIER GMBH
P.O. Box 1420
D-7990 Friedrichshafen 1
WEST GERMANY

Prof. Edward Haertel
School of Education
Stanford University
Stanford, CA 94305

Dr. Ronald K. Hambleton
University of Massachusetts
Laboratory of Psychometric
and Evaluative Research
Hills South, Room 152
Amherst, MA 01003

Dr. Delwyn Harnisch
University of Illinois
51 Gerty Drive
Champaign, IL 61820

Dr. Grant Henning
Senior Research Scientist
Division of Measurement
Research and Services
Educational Testing Service
Princeton, NJ 08541

Ms. Rebecca Hetter
Navy Personnel R&D Center
Code 63
San Diego, CA 92152-6800

Dr. Thomas M. Hirsch
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. Paul W. Holland
Educational Testing Service,
21-T
Rosedale Road
Princeton, NJ 08541

Dr. Paul Horst
677 G Street, #184
Chula Vista, CA 92010

Ms. Julia S. Hough
Cambridge University Press
40 West 20th Street
New York, NY 10011

Dr. William Howell
Chief Scientist
AFHRL/CA
Brooks AFB, TX 78235-5601

Dr. Lloyd Humphreys
University of Illinois
Department of Psychology
603 East Daniel Street
Champaign, IL 61820

Dr. Steven Hunka
3-104 Educ. N.
University of Alberta
Edmonton, Alberta
CANADA T6G 2G5

Dr. Huynh Huynh
College of Education
Univ. of South Carolina
Columbia, SC 29208

Dr. Robert Jannarone
Elec. and Computer Eng. Dept.
University of South Carolina
Columbia, SC 29208

Dr. Kumar Joag-dev
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright Street
Champaign, IL 61820

Dr. Douglas H. Jones
1280 Woodfern Court
Toms River, NJ 08753

Dr. Brian Junker
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Dr. Michael Kaplan
Office of Basic Research
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

Dr. Milton S. Katz
European Science Coordination Office
U.S. Army Research Institute
Box 65
FPO New York 09510-1500

Prof. John A. Keats
Department of Psychology
University of Newcastle
N.S.W. 2308
AUSTRALIA

Dr. Jwa-keun Kim
Department of Psychology
Middle Tennessee State University
P.O. Box 522
Murfreesboro, TN 37132

Mr. Soon-Hoon Kim
Computer-based Education Research Laboratory
University of Illinois
Urbana, IL 61801

Dr. G. Gage Kingsbury
Portland Public Schools
Research and Evaluation Department
501 North Dixon Street
P. O. Box 3107
Portland, OR 97209-3107

Dr. William Koch
Box 7246, Meas. and Eval. Ctr.
University of Texas-Austin
Austin, TX 78703

Dr. Richard J. Koubek
Department of Biomedical & Human Factors
139 Engineering & Math Bldg.
Wright State University
Dayton, OH 45435

Dr. Leonard Kroeker
Navy Personnel R&D Center
Code 62
San Diego, CA 92152-6800

Dr. Jerry Lehnus
Defense Manpower Data Center
Suite 400
1600 Wilson Blvd
Rosslyn, VA 22209

Dr. Thomas Leonard
University of Wisconsin
Department of Statistics
1210 West Dayton Street
Madison, WI 53705

Dr. Michael Levine
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Charles Lewis
Educational Testing Service
Princeton, NJ 08541-0001

Mr. Rodney Lim
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Dr. Robert L. Linn
Campus Box 249
University of Colorado
Boulder, CO 80309-0249

Dr. Robert Lockman
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. Frederic M. Lord
Educational Testing Service
Princeton, NJ 08541

Dr. Richard Luecht
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. George B. Macready
Department of Measurement Statistics & Evaluation
College of Education
University of Maryland
College Park, MD 20742

Dr. Gary Marco
Stop 31-E
Educational Testing Service
Princeton, NJ 08451

Dr. Clessen J. Martin
Office of Chief of Naval Operations (OP 13 F)
Navy Annex, Room 2832
Washington, DC 20350

Dr. James R. McBride
The Psychological Corporation
1250 Sixth Avenue
San Diego, CA 92101

Dr. Clarence C. McCormick
HQ, USMEPCOM/MEPCT
2500 Green Bay Road
North Chicago, IL 60064

Mr. Christopher McCusker
University of Illinois
Department of Psychology
603 E. Daniel St.
Champaign, IL 61820

Dr. Robert McKinley
Educational Testing Service
Princeton, NJ 08541

Mr. Alan Mead
c/o Dr. Michael Levine
Educational Psychology
210 Education Bldg.
University of Illinois
Champaign, IL 61801

Dr. Timothy Miller
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. Robert Mislevy
Educational Testing Service
Princeton, NJ 08541

Dr. William Montague
NPRDC Code 13
San Diego, CA 92152-6800

Ms. Kathleen Moreno
Navy Personnel R&D Center
Code 62
San Diego, CA 92152-6800

Headquarters Marine Corps
Code MPI-20
Washington, DC 20380

Dr. Ratna Nandakumar
Educational Studies
Willard Hall, Room 213E
University of Delaware
Newark, DE 19716

Library, NPRDC
Code P201L
San Diego, CA 92152-6800

Librarian
Naval Center for Applied Research in Artificial Intelligence
Naval Research Laboratory
Code 5510
Washington, DC 20375-5000

5/1/90

Dr. Harold F. O'Neil, Jr.
School of Education - WPH 801
Department of Educational
Psychology & Technology
University of Southern
California
Los Angeles, CA 90089-0031

Dr. James B. Olsen
WICAT Systems
1875 South State Street
Orem, UT 84058

Office of Naval Research,
Code 1142CS
800 N. Quincy Street
Arlington, VA 22217-5000
(6 Copies)

Dr. Judith Orasanu
Basic Research Office
Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard St.
Alexandria, VA 22311

Dr. Peter J. Pashley
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Wayne M. Patience
American Council on Education
GED Testing Service, Suite 20
One Dupont Circle, NW
Washington, DC 20036

Dr. James Paulson
Department of Psychology
Portland State University
P.O. Box 751
Portland, OR 97207

Dept. of Administrative Sciences
Code 54
Naval Postgraduate School
Monterey, CA 93943-5026

Dr. Mark D. Reckase
ACT
P. O. Box 168
Iowa City, IA 52243

Dr. Malcolm Ree
AFHRL/MOA
Brooks AFB, TX 78235

Mr. Steve Reiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455-0344

Dr. Carl Ross
CNET-PDCD
Building 90
Great Lakes NTC, IL 60088

Dr. J. Ryan
Department of Education
University of South Carolina
Columbia, SC 29208

Dr. Fumiko Samejima
Department of Psychology
University of Tennessee
310B Austin Peay Bldg.
Knoxville, TN 37916-0900

Mr. Drew Sands
NPRDC Code 62
San Diego, CA 92152-6800

Lowell Schoer
Psychological & Quantitative
Foundations
College of Education
University of Iowa
Iowa City, IA 52242

Dr. Mary Schratz
905 Orchid Way
Carlsbad, CA 92009

Dr. Dan Segall
Navy Personnel R&D Center
San Diego, CA 92152

Dr. Robin Shealy
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Dr. Kazuo Shigemasu
7-9-24 Kugenuma-Kaigan
Fujisawa 251
JAPAN

Dr. Randall Shumaker
Naval Research Laboratory
Code 5510
4555 Overlook Avenue, S.W.
Washington, DC 20375-5000

Dr. Richard E. Snow
School of Education
Stanford University
Stanford, CA 94305

Dr. Richard C. Sorensen
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Judy Spray
ACT
P.O. Box 168
Iowa City, IA 52243

Dr. Martha Stocking
Educational Testing Service
Princeton, NJ 08541

Dr. Peter Stoloff
Center for Naval Analysis
4401 Ford Avenue
P.O. Box 16268
Alexandria, VA 22302-0268

Dr. William Stout
University of Illinois
Department of Statistics
101 Illini Hall
725 South Wright St.
Champaign, IL 61820

Dr. Hariharan Swaminathan
Laboratory of Psychometric and
Evaluation Research
School of Education
University of Massachusetts
Amherst, MA 01003

Mr. Brad Sympson
Navy Personnel R&D Center
Code-62
San Diego, CA 92152-6800

Dr. John Tangney
AFOSR/NL, Bldg. 410
Bolling AFB, DC 20332-6448

Dr. Kikumi Tatsuoka
Educational Testing Service
Mail Stop 03-T
Princeton, NJ 08541

Dr. Maurice Tatsuoka
220 Education Bldg
1310 S. Sixth St.
Champaign, IL 61820

Dr. David Thissen
Department of Psychology
University of Kansas
Lawrence, KS 66044

Mr. Thomas J. Thomas
Johns Hopkins University
Department of Psychology
Charles & 34th Street
Baltimore, MD 21218

Mr. Gary Thomasson
University of Illinois
Educational Psychology
Champaign, IL 61820

Dr. Robert Tsutakawa
University of Missouri
Department of Statistics
222 Math. Sciences Bldg.
Columbia, MO 65211

Dr. Ledyard Tucker
University of Illinois
Department of Psychology
603 E. Daniel Street
Champaign, IL 61820

Dr. David Vale
Assessment Systems Corp.
2233 University Avenue
Suite 440
St. Paul, MN 55114

Dr. Frank L. Vicino
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Howard Wainer
Educational Testing Service
Princeton, NJ 08541

Dr. Michael T. Waller
University of Wisconsin-Milwaukee
Educational Psychology Department
Box 413
Milwaukee, WI 53201

Dr. Ning-Hai Wang
Educational Testing Service
Mail Stop 03-T
Princeton, NJ 08541

Dr. Thomas A. Warm
FAA Academy AAC934D
P.O. Box 25082
Oklahoma City, OK 73125

Dr. Brian Waters
HumRRO
1100 S. Washington
Alexandria, VA 22314

Dr. David J. Weiss
N660 Elliott Hall
University of Minnesota
75 E. River Road
Minneapolis, MN 55455-0344

Dr. Ronald A. Weitzman
Box 146
Carmel, CA 93921

5/1/90

Major John Welsh
AFHRL/MOAN
Brooks AFB, TX 78223

Dr. Douglas Wetzel
Code 51
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. Rand R. Wilcox
University of Southern California
Department of Psychology
Los Angeles, CA 90089-1061

German Military Representative
ATTN: Wolfgang Wildgrube
Streitkraefteamt
D-5300 Bonn 2
4000 Brandywine Street, NW
Washington, DC 20016

Dr. Bruce Williams
Department of Educational Psychology
University of Illinois
Urbana, IL 61801

Dr. Hilda Wing
Federal Aviation Administration
800 Independence Ave, SW
Washington, DC 20591

Mr. John H. Wolfe
Navy Personnel R&D Center
San Diego, CA 92152-6800

Dr. George Wong
Biostatistics Laboratory
Memorial Sloan-Kettering Cancer Center
1275 York Avenue
New York, NY 10021

Dr. Wallace Wulfeck, III
Navy Personnel R&D Center
Code 51
San Diego, CA 92152-6800

Dr. Kentaro Yamamoto
02-T
Educational Testing Service
Rosedale Road
Princeton, NJ 08541

Dr. Wendy Yen
CTB/McGraw Hill
Del Monte Research Park
Monterey, CA 93940

Dr. Joseph L. Young
National Science Foundation
Room 320
1800 G Street, N.W.
Washington, DC 20550

Mr. Anthony R. Zara
National Council of State Boards of Nursing, Inc.
625 North Michigan Avenue
Suite 1544
Chicago, IL 60611